MAINE EDUCATIONAL ASSESSMENT

2000–2001
Technical Manual

TABLE OF CONTENTS

CHAPTER 1—BACKGROUND AND OVERVIEW

PURPOSE OF THIS MANUAL

The purpose of this technical manual is to document the technical aspects of the 2000–2001 Maine Educational Assessment (MEA). In the fall of 2000, students in grades 4, 8, and 11 participated in the administration of the revised MEA in writing, reading, and health education. In the spring of 2001, students in grades 4, 8, and 11 were administered tests in mathematics, science and technology, social studies, and visual and performing arts. This report provides information about the technical quality of those assessments, including a description of the processes used to develop, administer, and score the tests and to analyze the test results. This report is intended to serve as a guide for replicating and/or improving the procedures in subsequent years.

While some parts of this technical report may be used by educated laypersons, the intended audience is experts in psychometrics and educational research. The report assumes a working knowledge of measurement concepts such as "reliability" and "validity," and statistical concepts such as "correlation" and "central tendency." In some chapters, the reader is presumed also to have basic familiarity with advanced topics in measurement and statistics.

LEARNING RESULTS

Following enactment of the Education Reform Act of 1984, Maine schools undertook a wide variety of initiatives designed to improve the quality of teaching and learning. Many of the lessons learned from those initiatives informed *Maine's Common Core of Learning*, a document published in 1990 that articulates a common vision for education in Maine by defining the knowledge, skills, and attitudes that all students should possess upon graduation from high school. In 1993, the Legislature directed the State Board of Education to undertake the next step in education reform by establishing a Task Force on *Learning Results* that was directed to

"develop long-range education goals and standards for school performance and student performance to improve learning results and recommend to the commissioner and to the Legislature a plan for achieving those goals and standards."

After substantial work, in January of 1996 the Task Force presented to the Legislature a report that contained a series of recommendations together with a set of standards, a plan for implementation, and proposed legislation. After a series of intense hearings during the 1996 Legislative Session, the Legislature adopted much of the work of the Task Force and directed the Department of Education and the State Board of Education to continue to develop the *Learning Results.*

Acting on the recommendations of the Task Force, the Legislature adopted six Guiding Principles that describe the characteristics of a well-educated person. To fulfill these principles, the Legislature required that the Department of Education and the State Board of Education develop *Learning Results* within the following eight areas:

Career Preparation
English Language Arts
Health and Physical Education
Mathematics
Modern and Classical Languages
Science and Technology
Social Studies
Visual and Performing Arts

These are not "subjects" in the same sense that we use the word when referring to courses in school. They are areas of learning that will in some cases cut across a number of discrete courses or disciplines. In response to the legislative directive, the Commissioner appointed a working group, known as the Critical Review Committee, to prepare a draft of standards for consideration by the State Board of Education and by the Legislature. The Committee met on numerous occasions during the summer and fall of 1996 to produce this revised document, which was approved in May of 1997 by the 118th Legislature.

PURPOSES OF THE MEA

The *Learning Results* are just one part of an educational system. As goals for what all students should know and be able to do upon finishing school, they are not written to prescribe a minimum of "passing" standard. The

setting of minimum requirements is the function of assessments that are separate from the creation of academic goals.

Because some students are ready for assessment at earlier stages than others, no assumption is made about when a standard might be achieved.

"The statute passed in April of 1996 includes the following provisions relating to assessment:

Student achievement of the learning results. . .must be measured by a combination of state and local assessments to measure progress and ensure accountability. The 4th-grade, 8th-grade, and 11th-grade results of the Maine Educational Assessment, the "MEA," are the state assessments used to measure achievement of the learning results. The 4th-grade and 8th-grade MEA must be used to measure achievement of the learning results beginning in the 1998-99 school year. Local school administrative units may develop additional assessments to measure achievement of the learning results, including student portfolios, performances, demonstrations, and other records of achievements."

An Assessment Design Team comprised of Maine educators and assessment specialists has been established to redesign state level assessments and to assist in development of high-quality local assessments that will be used to measure student achievement of the *Learning Results*. The statewide assessment system they are developing will

- align with Maine's *Learning Results;*

- utilize multiple measures of learning;

- ensure fair and equitable assessment for all students;

- utilize recognized, relevant technical standards for assessment;

- provide understandable information to educators, parents, students, the public, and the media;

- provide professional development opportunities for teachers, administrators, and future educators; and

- be practical and manageable.

ORGANIZATION OF THIS MANUAL

The organization of this manual is based on the conceptual flow of an assessment's life span; it begins with the initial test specification and addresses all the intermediate steps that lead to final score reporting. Section I covers

the development of the MEA tests. It consists of eight chapters, covering general design issues, the test development process, and the specific designs of the English language arts, mathematics, science and technology, social studies, visual and performing arts, and health education assessments. Section II consists of a single chapter describing the administration of the tests. Section III contains six chapters covering scoring, equating and scaling, item analysis, reliability, validity, and reporting. We have also included two additional sections: Section IV contains references and Section V contains the appendices.

SECTION I: ASSESSMENT DEVELOPMENT

CHAPTER 2—OVERVIEW OF TEST DESIGN

LEARNING RESULTS

MEA questions are directly linked to the **content standards** and **performance indicators** described in Maine's *Learning Results*. The content standards are the basis for the reporting categories developed for each subject area; the performance indicators are used to help guide the development of test questions. No other content or process is subject to statewide assessment. An item may address part, all, or several of the performance indicators.

ITEM TYPES

Maine's educators and students were familiar with the question types that were used in the 2000-01 assessment program as all had been previously introduced. The types of questions used and the functions of each are described below.

Multiple-choice questions were used, in part, to provide breadth of coverage of a subject area. Because they require no more than a minute for most students to answer, these questions make efficient use of limited testing time and allow coverage of a wide range of knowledge and skills.

Short-answer questions were used to assess students' skills and their abilities to work with brief, well-structured problems that had one or a very limited number of solutions (e.g., mathematical computations). Short-answer questions require approximately two to five minutes for most students to answer. The advantage of this type of question is that it requires students to demonstrate knowledge and skills by generating, rather than merely selecting, an answer.

Constructed-response questions typically require students to use higher-order thinking skills—evaluation, analysis, summarization, and so on—in constructing a satisfactory response. Constructed-response questions should take most students approximately five to ten minutes to complete. It should be noted that the use of previously released MEA questions to prepare students to answer this kind of question was appropriate and encouraged.

Extended-response questions assess students ability to analyze and solve challenging problems based on real-world, age-appropriate situations that call for multiple approaches and may have more than one solution. An ability to communicate and justify a solution through the use of writing, tables, charts, and/or graphic displays contributes to a student's success in many of the extended-response questions. This type of question requires approximately ten to twenty minutes for most students to complete.

COMMON-MATRIX DESIGN

The 2000-01 MEA continued to measure what students know and are able to do by using a variety of question types. The tests continued to be structured using both *common* and *matrix-sampled* questions. Common questions are those taken by all students at a given grade level; in addition, a larger pool of matrix-sampled questions is divided among the multiple forms of the test at each grade level. (There were 12 forms of the test in 2000-01.) Each student took only one form of the test and so answered a fraction of the matrix-sampled questions in the entire pool. This design, which has been used throughout the MEA's history, provides reliable and valid results at the student level. It also provides for a greater breadth of coverage of a subject area for school results while minimizing testing time.

In 2000–01, the reports continued to report out only common scores in the results for ease of understanding them. If student results were based on common and matrix-sampled questions, one student could score higher than another in raw score, but lower in scaled score. By producing common results only, this type of reversal was avoided.

TEST SESSION TIMES

The MEA tests were given at two different times during the school year: **writing, reading, and health education** were administered to all grades in late fall, and tests in **mathematics, science and technology, social studies, and visual and performing arts** were administered to all grades during a two-week period in early March. Schools were able to schedule testing sessions at any time during the first week of this period, provided they followed the sequence in the scheduling guidelines detailed in test administration manuals. The second week was reserved for make-up testing of students who were absent from initial test sessions.

The timing and scheduling guidelines for MEA tests were based on estimates of the time it would take an average student to respond to each type of question that makes up the test:

- multiple-choice questions – 1 minute per question;

- short-answer questions – 2 minutes per question;

- constructed-response questions – 10 minutes per question;

- extended-response questions – 20 minutes per question.

For the English language arts reading test, the scheduling guidelines included an estimate of 10 minutes to read each passage used in the assessment.

While the guidelines for scheduling are based on the assumption that most students will complete the test within the time estimated, each test session was scheduled so that additional time was provided for students who needed it. One-third additional time was allocated for each session (i.e., 60-minute sessions were scheduled with an additional 20 minutes; 45-minute sessions with an additional 15 minutes; and 35-minute sessions with an additional 10 minutes).

If additional classroom space was not available for students who required additional time to complete the tests, schools were allowed to consider using another space, such as the guidance office, for this purpose. If additional areas were not available, it was recommended that each classroom being used for test administration be scheduled for the maximum amount of time. Detailed instructions on test administration and scheduling were provided in the coordinator's and administrator's manuals.

CHAPTER 3—TEST DEVELOPMENT PROCESS

DEVELOPMENT COMMITTEE ITEM IDEA GENERATION

The development of the MEA tests continues to be a cooperative effort by content development committees comprising Maine teachers, curriculum supervisors, higher education faculty, content specialists of the Department of Education, and curriculum/assessment specialists employed by the program's contractor, Measured Progress. The committees are structured to represent all areas of the state and committee members all serve rotating terms.

The committees' primary roles are to develop test questions for the MEA and to interpret testing data so that those questions could be selected for the program. The 2001-02 MEA development committee for each subject area at grade levels 4, 8, and 11 met several times. In the development phase, the committees reviewed the content standards and test specifications; then they brainstormed or drafted test questions and scoring rubrics to fit those specifications. After the questions were field tested, the committees reviewed the field-test data and made recommendations about selecting, revising, or eliminating specific questions from the item pool for the operational test. At that time, the committees also confirmed that each question conformed directly to Maine's *Learning Results* and was thus assigned to the appropriate content standard reported in school and district results. Because many MEA questions are released to the public each year, the committees repeat these activities annually as new questions are developed in order to replenish the item pool.

INTERNAL ITEM REVIEW

- The lead or peer test developer within the content specialty reviewed the typed item, constructed-response scoring guide, and any reading selections and graphics.

- The content reviewer considered item "integrity;" item content and structure; appropriateness to designated content area; item format; clarity; possible ambiguity; keyability; single "keyness;" appropriateness and quality of reading selections and graphics; and appropriateness of scoring guide descriptions and distinctions (as correlated to the item and within the guide itself).

- The content reviewer also considered scorability and evaluated whether the scoring guide adequately addressed performance on the item.

- Fundamental questions the content reviewer considered, but was not limited to, included the following:

    – What is the item asking?

    – Is the key the only possible key?

    – Is the constructed-response item scorable as written (were the correct words used to elicit the response defined by the guide)?

    – Is the wording of the scoring guide appropriate and parallel to the item wording?

    – Is the item complete (e.g., with scoring guide, content codes, key, grade level, and contract identified)?

    – Is the item appropriate for the designated grade level?

EXTERNAL ITEM REVIEW

- Item sets were brought to Content Development Committee meetings for review and revision.

ITEM EDITING

Editors reviewed and edited the items from the Content Development Committee item review to ensure uniform style (based on *The Chicago Manual of Style, 14th Edition*) and adherence to sound testing principals. These principles included the stipulation that items

- were correct with regard to grammar, punctuation, usage, and spelling;

- were written in a clear, concise style;

- contained unambiguous explanations to students as to what is required to attain a maximum score;

- were written at a reading level that would allow the student to demonstrate his or her knowledge of the tested subject matter, regardless of reading ability;

- exhibited high technical quality regarding psychometric characteristics;

- had appropriate answer options or score-point descriptors; and

- were free of potentially sensitive content.

REVIEWING AND REFINING

Test developers presented item statistics to the development committees to assist in the committees'
recommendations for placement of items into the common and matrix portions of the test. The Department of
Education made the final selections with the assistance of Measured Progress at a meeting.

OPERATIONAL TEST ASSEMBLY

Test assembly is the sorting and laying out of item sets into test forms. Criteria considered during this process
included the following:

- **Content coverage/match to test design.** The curriculum specialist completed an initial sorting of items into
  sets based on a balance of content categories across sessions and forms, as well as a match to the test design
  (e.g., number of multiple-choice, short-answer, constructed-response, and extended-response items).

- **Item difficulty and complexity.** Item statistics drawn from the data analysis of previously tested items were
  used to ensure that there were similar levels of difficulty and complexity across forms.

- **Visual balance.** Item sets were reviewed to ensure that each reflected a similar length and "density" of
  selected items (e.g., length/complexity of reading selections, or number of graphics).

- **Option balance.** Each item set was checked to verify that it contained a roughly equivalent number of key
  options (As, Bs, Cs, and Ds).

- **Name balance.** Item sets were reviewed to ensure that a diversity of names was used.

- **Bias.** Each item set was reviewed to ensure fairness and balance based on gender, ethnicity, religion, socio-
  economic status, and other factors.

- **Page fit.** Item placement was modified to ensure the best fit and arrangement of items on any given page.

- **Facing page issues.** For multiple items associated with a single stimulus (a graphic or reading selection),
  consideration was given to whether those items needed to begin on a left- or right-hand page, as well as to
  the nature and amount of material that needed to be placed on facing pages. These considerations served to
  minimize the amount of "page flipping" required of the students.

- **Relationships between forms.** Sets of common items were placed identically in each version of the forms.
  Although matrix-sampled item sets differ from form to form, they must take up the same number of pages in

each form so that sessions and content areas begin on the same page in every form. Therefore, the number of pages needed for the longest form often determines the layout of each form.

- **Visual appeal.** The visual accessibility of each page of the form was always taken into consideration, including such aspects as the amount of "white space," the density of the text, and the number of graphics.

EDITING DRAFTS OF OPERATIONAL TESTS

Any changes made by the test construction specialist must be reviewed and approved by the test developer. Once a form had been laid out in what was considered its final form, it was reread to identify any final considerations, including the following:

- Editorial changes. All text was scrutinized for editorial accuracy, including consistency of instructional language, grammar, spelling, punctuation, and layout. Measured Progress' publishing standards are based on *The Chicago Manual of Style, 14th Edition*.

- "Keying" items. Items were reviewed for any information that might "key" or provide information that would help answer another item. Decisions about moving keying items are based on the severity of the "key-in" and the placement of the items in relation to each other within the form.

- Key patterns. The final sequence of keys was reviewed to ensure that their order appeared random (e.g., no recognizable pattern, and no more than three of the same key in a row).

BRAILLE AND LARGE-PRINT TRANSLATION

Form 1 for grades 4, 8, and 11 tests was translated into Braille by a subcontractor that specializes in test materials for blind and visually impaired students. In addition, Form 1 for each grade was adapted into a large-print version.

CHAPTER 4—DESIGN OF ENGLISH LANGUAGE ARTS ASSESSMENT

READING

BLUEPRINT

As indicated earlier, the English language arts framework for reading is based on Maine's *Learning Results*, which identifies five **content standards** that apply specifically to reading and reading comprehension. Those content standards are:

- **Process of reading:** Students use the skills and strategies of the reading process to comprehend, interpret, evaluate, and appreciate what they have read.

- **Literature and culture:** Students use reading, listening, and viewing strategies to experience, understand, and appreciate literature and culture.

- **Language and images:** Students demonstrate an understanding of how words and images communicate.

- **Informational texts:** Students apply reading, listening, and viewing strategies to informational texts across all areas of curriculum.

The content standards have been adapted to create a reporting category framework for reading, as shown below.

Passage Type	Comprehension of Literary and Informational Texts			Total
	Reading Comprehension and Literary Analysis	A. Process of Reading	C. Language and Images	
B. Literature and Culture: Literary Passages				50%
D. Informational Texts: Content Passages Practical Passages				50% (30%) (20%)
Total	80%	20%		100%

CONTENT SPECIFICATIONS

The first major reporting category at the student, school, and district levels is "comprehension of literary and informational texts." The data generated for this reporting category was based on questions related to three types of reading passages that reflect standards B and D of the English Language Arts (ELA) *Learning Results*. The passage

types were identical to those that have been used in the MEA in past years. Fifty percent of the passages comprised literary works; 30% were selected from content pieces (see explanation below); and 20% were drawn from practical sources (see explanation below).

Passages included both long and short "authentic" texts selected from reading sources that students at each grade level would be likely to encounter in their classroom and in their independent reading. The passages were not written specifically for the assessment, but instead were collected from published works.

- **Literary passages** are represented by a variety of genres—modern narratives; diary entries; drama; poetry; biographies; essays; excerpts from novels; short stories; and traditional narratives, such as fables, myths, and folktales.

- **Content passages** are primarily informational and often deal with the areas of science and social studies. They are drawn from such sources as newspapers, magazines, and books.

- **Practical passages** are functional materials that instruct or advise the reader—for example, directions, reference tools, or manuals.

The main difference in the passages used for grades 4, 8, and 11 is the degree of difficulty. All passages were selected to be appropriate for the intended audience; however, the ideas expressed become increasingly more complex at grade levels 8 and 11.

The questions related to these passages require students to demonstrate their skills in both literal comprehension (where the answer is stated explicitly in the text) and inferential comprehension (where the answer is implied by the text and/or the text must be connected to relevant prior knowledge to determine an answer). In addition, some questions focus on the reading skills reflected in content standards A and C of the *Learning Results*. Questions of this type require students to use the skills and strategies of reading to answer questions—for example, how to identify the author's principal purpose, such as to persuade, entertain, or inform—and to demonstrate their understanding of how words and images communicate to readers.

ITEM TYPES

The MEA English language arts assessment in reading included multiple-choice, short-answer, and constructed-response questions, as well as one extended-response/writing sample question. Short-answer questions,

which were new in the revised MEA, required students to write an answer consisting of several phrases or short sentences. Each type of question was worth a specific number of points in the student's total language arts score, as shown below.

Type of Question	Possible Score Points
Multiple Choice	0–1
Short Answer	0–2
Constructed Response	0–4
Extended Response/Writing Sample	0–8

TEST DESIGN

The table below summarizes the numbers and associated questions that were used in the MEA reading assessment for 2000-01.

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
2A	6	2	1	0					25 (+10)
2B	6	2	2	0					25 (+10)
3A	6	1	0	1					45 (+15)
3B					6	2	1	0	25 (+10)

Key
- MC = multiple-choice questions
- SA = short-answer questions
- CR = constructed-response questions
- ER = extended-response/writing sample question

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

ENGLISH LANGUAGE ARTS — READING
GRADE 4

	Standard A				Standard B				Standard C				Standard D				Total Points
	MC x1	SA x2	CR x4	ER x8	MC x1	SA x2	CR x4	ER x8	MC x1	SA x2	CR x4	ER x8	MC x1	SA x2	CR x4	ER x8	216
Common Passages																	
Home Grown Hydras	1												2	1	1		9
On My Own With Alex	3									1			3	1	1		14
Rats Don't	1				3	1		1	2								16
Spring Poems	1				1		1		1	1							9
Matrix Passages																	
Alcove Spring	2				4	1	1			1							14
Amazing Spiders	1								2					1			5
Avalanche!	2	1			4	1	1										14
Bacon-Tomato Sandwiches	1		1										2	1			9
Be a Junk Food Detective	1												2	1			5
Brian's Winter	1				1	1			1								5
Cleaning Up the Ocean	1	1											2				5
Climbing/Every Time I Climb			1		1	1			2								9
Drinking Milk is Good for Birds	1								1				4	2	1		14
I'm Going to Be Famous	2				3	2	1		1								14
Let's Write a True Life Story	1								1				1	1	1		9
Marsha	1					1	1		2								9
My Dino Discovery									1	1			2		1		9
One Brave Summer					1	1	1		2								9
Ruby	1				1	1			1								5
Secret Place					2	1			1								5
Should Schools Tell You...	3				1		1						2	2			14
Welcome to the Inventors Club	2	1							1				3	1	1		14

ENGLISH LANGUAGE ARTS — READING
GRADE 8

	Standard A				Standard B				Standard C				Standard D				Total
	MC x1	SA x2	CR x4	ER x8	MC x1	SA x2	CR x4	ER x8	MC x1	SA x2	CR x4	ER x8	MC x1	SA x2	CR x4	ER x8	Points 216
Common Passages																	
An American Childhood					4	1		1	2								16
Principles of Art									1		1		2	1			9
The Base Stealer					1	1	1		2								9
The Debate Over Closing…	1	1							1				4	1	1		14
Matrix Passages																	
Amir					6	2					1						14
Children of the River		1			3												5
Cool Science-A lesson runs…									1	1			2				5
First Lesson/Fathers			1		2	1			1								9
Gentle friends/essential allies													3	1			5
Go Fly a Kite	1								1				4	2	1		14
Graduation Morning					2				1	1							5
Hurricanes									3				3	2	1		14
Lost in the Woods	2												4	2	1		14
Niagara Falls	1												2	1	1		9
Right Smart O' Wind	2		1		1	2			3								14
Road Runner Sports	2												1	1	1		9
The Life of the Ladybird Beetle			1						1	1			5	1			14
Uncle Joe	2	1			4	1	1										14
Why I Never Shoot Bears	1				1	1	1		1								9
Wreck of the Monkey Cage	1				1	1	1		1								9
You Can Be An Inventor	2								1					1			5

ENGLISH LANGUAGE ARTS — READING
GRADE 11

	Standard A				Standard B				Standard C				Standard D				Total Points
	MC x 1	SA x 2	CR x 4	ER x 8	MC x 1	SA x 2	CR x 4	ER x 8	MC x 1	SA x 2	CR x 4	ER x 8	MC x 1	SA x 2	CR x 4	ER x 8	216
Common Passages																	
Deer Among Cattle					1	1	1		2								9
Mt. Katahdin Via the Knife Edge	1				1	1	1		1								9
New Directions	1	1			4	1	1		1								14
Sweet Season	1								2				3	1		1	16
Matrix Passages																	
A Day at the Theatre	1								1				4	2	1		14
A Presidential Candidate	1	1	1		4	1			1								14
At Harvesttime	1				1	1			1		1						9
Chief Joseph of the Nez Perce…	1												2	1	1		9
Children of the Sun	1				2	1											5
Dead Snails Leave No Trails	2								1				3	2	1		14
Discover White Water Rafting	1												2	1			5
Feet	1						1		1				1	1			9
I Wandered Lonely As a Cloud					1	1			2								5
Life in the Thirteen Colonies	1						1										5
Nearer	1				2								2		1		9
Piltdown Man	1								2				2				5
Polonius's Advice to Laertes	2				1	1	1										9
Prevent Repetitive Strain At…	1	1											2		1		9
Reading Moving Water	1								1				4	2	1		14
The House on Mango Street	1				2	1											5
The Ojibwa Corn Hero	1	1			5	1	1										14
William	1				4	2	1		1								14

WRITING

BLUEPRINT

The MEA assesses students' writing skills directly through the use of writing prompts, or topics, to which students respond. Maine's *Learning Results* includes two content standards that apply specifically to writing. Those content standards are

- **Standard English conventions:** Students write and speak correctly, using conventions of standard written and spoken English.

- **Stylistic and rhetorical aspects of writing and speaking:** Students use stylistic and rhetorical aspects of writing and speaking to explore ideas, to present lines of thought, to represent and reflect on human experience, and to communicate feelings, knowledge, and opinions.

Note: Standard E, processes of writing and speaking, addressed students' abilities to use the skills and strategies of the writing process. This standard was assessed at the local level only.

The *Learning Results* standards were adapted to create reporting categories for writing, as shown below.

Stylistic and Rhetorical Aspects of Writing	▪ Idea/topic development ▪ Organization ▪ Supporting detail
Standard English Conventions	▪ Grammar ▪ Spelling ▪ Punctuation ▪ Capitalization ▪ Sentence structure

CONTENT SPECIFICATIONS

Four broad types, or modes, of writing are used in the MEA, as listed below[1]:

- **Narration:** Narrative writing answers the question, "What happened?" It tells a story through a sequence of events, so that the reader understands the action.

- **Exposition:** Expository writing informs the reader about something. Methods of exposition include comparison and contrast, illustration, classification, definition, and analysis. Methods of exposition are often combined to accomplish a specific purpose for writing.

---

[1] Descriptions are adapted from *Modern Rhetoric*, by Cleanth Brooks and Robert Penn Warren.

- **Description:** Descriptive writing presents the qualities of objects, persons, conditions, and actions.

- **Persuasion/argument:** Persuasive writing uses emotional appeals to bring about a change of attitude, point of view, or feeling. Argumentative writing uses logic and reason to bring about a change of attitude, point of view, or feeling; it shows that a conclusion merits belief because of credible data, evidence, and so on.

The student's "audience" and "purpose for writing" also influence the development, style, and tone of a written composition. These were specified as part of the prompts and varied by grade level. In addition, the prompts were developed with the following criteria as guidelines:

- the prompts must be interesting to students;

- the prompts must be accessible to all students (i.e., all students would have something to say about the topic); and

- the prompts must generate sufficient text to be effectively scored.

The prompts used in the 2000-01 MEA writing assessment follow.

Grade 4 prompt: You find something special. Describe what it is and what you do with it.

Grade 8 prompt: Write a letter to a student who is about to enroll in your school. Tell this new student what he or she should expect.

Grade 11 prompt: Explain how high school students view a particular aspect or issue of life differently than adults view it.

TEST DESIGN

Each student responded to one common writing prompt, as well as a common extended-response question that was scored for both reading and writing. The chart below outlines the total number of possible points—as reported—by learning results and item type.

ENGLISH LANGUAGE ARTS—WRITING
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common Prompt	Extended Response Writing	Total Points
Standard English Conventions (Standard F)	8	4	12
Stylistic and Rhetorical Aspects of Writing (Standard G)	12	6	18

**Number of Points Possible
Grade 8**

Standard	Common Prompt	Extended Response Writing	Total Points
Standard English Conventions (Standard F)	8	4	12
Stylistic and Rhetorical Aspects of Writing (Standard G)	12	6	18

**Number of Points Possible
Grade 11**

Standard	Common Prompt	Extended Response Writing	Total Points
Standard English Conventions (Standard F)	8	4	12
Stylistic and Rhetorical Aspects of Writing (Standard G)	12	6	18

CHAPTER 5—DESIGN OF THE MATHEMATICS ASSESSMENT

BLUEPRINT

The mathematics framework was based on Maine's *Learning Results*, which identifies eleven **content standards** as shown below:

- **Numbers and number sense:** Students understand and demonstrate a sense of what numbers mean and how they are used.

- **Computation:** Students understand and demonstrate computation skills.

- **Data analysis and statistics:** Students understand and apply concepts of data analysis.

- **Probability:** Students understand and apply concepts of probability.

- **Geometry:** Students understand and apply concepts from geometry.

- **Measurement:** Students understand and demonstrate measurement skills.

- **Patterns, relations, and functions:** Students understand that mathematics is the science of patterns, relationships, and functions.

- **Algebra concepts:** Students understand and apply algebraic concepts.

- **Discrete mathematics:** Students understand and apply concepts in discrete mathematics.

- **Mathematical reasoning:** Students understand and apply concepts of mathematical reasoning.

- **Mathematical communication:** Students reflect upon and clarify their understanding of mathematical ideas and relationships.

These standards were used to create a reporting category framework for mathematics, shown below. The framework was divided into two major areas:

- **content**, which refers to the student's knowledge and conceptual and procedural understanding of each standard, and

- **application**, which refers to a student's use of knowledge and conceptual and procedural understanding as a basis for application through reasoning, inquiry, communication of ideas, and problem solving.

Each question in the mathematics assessment measured a content standard; in addition, each question was reported as measuring either content or application.

As shown in the table below, the goal for distribution of questions, or emphasis, across standards varies from grade to grade.

Content Standard	Grade		
	4	8	11
A. Number and Number Sense	15%	14%	10%
B. Computation	15%	11%	5%
C. Data Analysis and Statistics	12%	11%	10%
D. Probability	8%	11%	10%
E. Geometry	12%	11%	15%
F. Measurement	12%	10%	10%
G. Patterns, Relations, Functions	12%	13%	15%
H. Algebra Concepts	9%	14%	15%
I. Discrete Mathematics	5%	5%	10%

CONTENT AND APPLICATION

For students to function effectively as mathematical problem-solvers, they must be taught how to apply and communicate basic concepts and procedures as well as how to do the procedures. **Content questions** measure what students have been taught directly, including the basic concepts and procedural skills from all the content standards. For example, in the numbers and number sense standard and the computation standard, conceptual and procedural knowledge includes understanding of place value in our number system; the computational algorithms as applied to whole numbers, fractions, and decimals; and the concepts of ratio, proportion, and percent. In the data analysis and statistics standard, conceptual and procedural knowledge includes the reading of charts and graphs as well as the concepts of averages (means, medians, and modes) and methods for computing them. Contextual settings used in questions measuring this category are very simple and are directly related to those used in the teaching of the concepts and procedures.

Application questions measure what the students can do with what they have been taught. Included are questions requiring students to combine the basic concepts and procedures to solve real-life and mathematical problems, to evaluate their own ideas and the ideas of others using mathematical reasoning, and to communicate their ideas using the wealth of symbolic, pictorial, graphic, and verbal representations available in mathematics.

It is important to understand that application questions also measure mastery of the basic concepts and procedures. For example, in mathematics, 20% of the questions are either constructed- or extended-response questions (see "Content Specifications" below), which are worth up to 4 and 8 score points respectively. In most cases, portions of these questions require the student to perform some problem solving, reasoning, and/or communicating, and so the questions are classified under applications. At the same time, however, the questions require students to demonstrate their understanding of mathematics content. If a student does not show mastery of all aspects of a constructed- or extended-response question, or if he/she makes careless errors, the student does not earn the highest score for that question. Thus, it can be said that **all** mathematics questions in the MEA measure content; some questions go beyond that realm, however, and are classified for reporting purposes as application.

CONTENT SPECIFICATIONS

The MEA mathematics assessment included multiple-choice, short-answer, constructed-response, and extended-response questions. Short-answer questions, which were new in the revised MEA, required students to perform a computation or solve a simple problem. Extended-response questions in mathematics are similar to constructed-response questions except that they are more complex, requiring 10 to 20 minutes of response time. Each type of question was worth a specific number of points in the student's total mathematics score, as shown below.

Type of Question	Possible Score Points
Multiple Choice	0–1
Short Answer	0–2
Constructed Response	0–4
Extended Response	0–8 (grades 8 and 11)

TEST DESIGN

The tables below summarize the numbers and types of questions that were used in the MEA mathematics assessment for 2000-01. The tables show the construction of the common and matrix-sampled portions of the assessment.

GRADE 4

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
4A (NC)	4	3	2	0	2	1	0	0	30 (+10)
4B (C)	8	1	1	0	2	0	1	0	30 (+10)
4C (C)	8	1	2	0	1	1	0	0	30 (+10)

GRADES 8 AND 11

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
4A NC	15	4	2	0	2	1	0	0	50 (+15)
4B (C)	5	1	1	1	3	1	1	0	50 (+15)

Key

- (C) = calculator use allowed
- (NC) = no calculator use allowed
- MC = multiple-choice questions
- SA = short-answer questions
- CR = constructed-response questions
- ER = extended-response questions

THE USE OF CALCULATORS IN THE MEA

The Maine educators who designed and developed the assessment test acknowledge the importance of mastering of arithmetic algorithms. At the same time, they understand that the use of calculators is a necessary and important skill in society today. Calculators can save time and error in the measurement of some higher order thinking skills and allow students to do more sophisticated and intricate problems. For these reasons, it was decided that calculators should be permitted in some parts of the MEA mathematics assessment and prohibited in others. (Students were allowed to use any calculator with which they are familiar.)

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

MATHEMATICS
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common				Matrix Per Form				Total Points 206
	MC x 1	SA x 2	CR x 4	Points	MC x 1	SA x 2	CR x 4	Points	
Content	15	4	2	31	51	7	0	65	96
Application	5	1	3	19	9	17	12	91	110
Numbers and Number Sense (Standard A)	8	0	1	12	9	2	1	17	29
Computation (Standard B)	0	1	1	6	7	7	1	25	31
Data Analysis and Statistics (Standard C)	0	0	1	4	13	1	1	19	23
Probability (Standard D)	0	0	1	4	6	1	1	12	16
Geometry (Standard E)	3	1	0	5	9	1	2	19	24
Measurement (Standard F)	4	1	0	6	7	3	2	21	27
Patterns, Relations, Functions (Standard G)	3	1	0	5	5	3	2	19	24
Algebra Concepts (Standard H)	0	1	1	6	3	3	1	13	19
Discrete Mathematics (Standard I)	2	0	0	2	1	3	1	11	13

Standard	Common					Matrix Per Form					Total Points 206
	MC x1	SA x2	CR x4	ER x8	Points	MC x1	SA x2	CR x4	ER x8	Points	
Content	16	4	2	0	32	45	14	3	0	85	117
Application	4	1	1	1	18	15	10	9	0	71	89
Numbers and Number Sense (Standard A)	3	1	1	0	9	9	3	1	0	19	28
Computation (Standard B)	5	0	0	0	5	8	3	1	0	18	23
Data Analysis and Statistics (Standard C)	3	1	0	0	5	8	3	1	0	18	23
Probability (Standard D)	1	0	1	0	5	6	2	2	0	18	23
Geometry (Standard E)	3	1	0	0	5	2	2	3	0	18	23
Measurement (Standard F)	0	2	0	0	4	3	3	2	0	17	21
Patterns, Relations, Functions (Standard G)	2	0	1	0	6	9	4	1	0	21	27
Algebra Concepts (Standard H)	0	0	0	1	8	10	3	1	0	20	28
Discrete Mathematics (Standard I)	3	0	0	0	3	5	1	0	0	7	10

MATHEMATICS
NUMBER OF POINTS POSSIBLE
GRADE 11

Standard	Common					Matrix Per Form					Total Points 206
	MC x1	SA x2	CR x4	ER x8	Points	MC x1	SA x2	CR x4	ER x8	Points	
Content	10	0	1	0	14	31	3	0	0	37	51
Application	10	5	2	1	36	29	21	12	0	119	155
Numbers and Number Sense (Standard A)	3	0	0	0	3	7	2	0	0	11	14
Computation (Standard B)	2	1	0	0	4	7	1	0	0	9	13
Data Analysis and Statistics (Standard C)	3	1	0	0	5	7	3	2	0	21	26
Probability (Standard D)	2	0	1	0	6	10	2	1	0	18	24
Geometry (Standard E)	3	1	1	1	17	8	3	2	0	22	39
Measurement (Standard F)	2	0	1	0	6	4	3	2	0	18	24
Patterns, Relations, Functions (Standard G)	2	1	0	0	4	8	3	2	0	22	26
Algebra Concepts (Standard H)	3	1	0	0	5	7	4	2	0	23	28
Discrete Mathematics (Standard I)	0	0	0	0	0	2	3	1	0	12	12

CHAPTER 6—DESIGN OF THE SCIENCE AND TECHNOLOGY ASSESSMENT

BLUEPRINT

The science and technology framework was based on Maine's *Learning Results*, which identify thirteen **content standards** as listed below:

- **Classifying life forms:** Students understand that there are similarities within the diversity of all living things.

- **Ecology:** Students understand how living things depend on one another and on non-living aspects of the environment.

- **Cells:** Students understand that cells are the basic units of life.

- **Continuity and change:** Students understand the basis for all life and that all living things change over time.

- **Structure of matter:** Students understand the structure of matter and the changes it can undergo.

- **The Earth:** Students gain knowledge about the Earth and the processes that change it.

- **The universe:** Students gain knowledge about the universe and how humans have learned about it, and the principles upon which it operates.

- **Energy:** Students understand concepts of energy.

- **Motion:** Students understand the motion of objects and how forces can change that motion.

- **Inquiry and problem solving:** Students apply inquiry and problem-solving approaches in science and technology.

- **Scientific reasoning:** Students learn to formulate and justify ideas and to make informed decisions.

- **Communication:** Students communicate effectively in the applications of science and technology.

- **Implications of science and technology:** Students understand the historical, social, economic, environmental, and ethical implications of science and technology.

Nine of these standards (A through I) address the various content areas in science and technology as shown below.

Content Standard	Grade		
	4	**8**	**11**
A. Classifying Life Forms	10%	10%	8%
B. Ecology	12%	10%	10%
C. Cells	10%	15%	12%
D. Continuity and Change	10%	10%	12%
E. Structure of Matter	8%	15%	15%
F. The Earth	10%	10%	15%
G. The Universe	15%	10%	8%
H. Energy	15%	10%	10%
I. Motion	10%	10%	10%

The remaining four (J, K, L, and M) highlight scientific applications. These have been adapted and combined to create the reporting category framework for science and technology, shown below.

Content Standard	Application			
	J. Inquiry and Problem Solving	K. Scientific Reasoning	L. Communication	M. Implications of Science & Technology
A. Classifying Life Forms				
B. Ecology				
C. Cells				
D. Continuity and Change				
E. Structure of Matter				
F. The Earth				
G. The Universe				
H. Energy				
I. Motion				

All questions in the science and technology assessment measured a content standard; approximately 40% of the questions were written to measure a performance indicator in applications.

APPLICATIONS

The score for applications refers to a student's use of knowledge and conceptual and procedural understandings as a basis for application through reasoning, inquiry, communication of ideas, and problem solving.

CONTENT SPECIFICATIONS

The MEA science and technology assessment included multiple-choice, short-answer, constructed-response, and extended-response questions. Short-answer questions, which were new in the revised MEA, required students to

formulate an answer using one or two words or a short phrase. Extended-response questions in science and technology are similar to constructed-response questions except that they are more complex, requiring 10 to 20 minutes of response time. Each type of question was worth a specific number of points in the student's total science and technology score, as shown below.

Type of Question	Possible Score Points
Multiple Choice	0–1
Short Answer	0–2
Constructed Response	0–4
Extended Response	0–8 (grades 8 and 11)

The scoring of extended response questions may utilize either two four-point guides, one measuring science content and one measuring science applications, or one eight-point guide, measuring solely content or applications.

TEST DESIGN

The tables below summarize the numbers and types of questions that were used in the MEA science and technology assessment for 2000-01.

GRADE 4

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
2A	7	1	2	0	2	1	0	0	30 (+10)
2B	7	2	1	0	2	0	1	0	30 (+10)
2C	6	2	2	0	2	0	0	0	30 (+10)

GRADES 8 AND 11

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
2A	13	1	1	1	2	1	0	0	50 (+15)
2B	7	4	2	0	4	0	1	0	50 (+15)

Key
- MC = multiple-choice questions
- SA = short-answer questions
- CR = constructed-response questions
- ER = extended-response questions

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

SCIENCE AND TECHNOLOGY
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common MC X1	SA X2	CR X4	Points	Matrix MC X1	SA X2	CR X4	Points	Total Possible Points
Content	12	3	3	30	44	7	7	86	116
Classifying Life Forms (Standard A)	1			1	4	1	2	14	15
Ecology (Standard B)	1			1	6		1	10	11
Cells (Standard C)	1		1	5	5	1	1	11	16
Continuity and Change (Standard D)	2	1		4	5	1		7	11
Structure of Matter (Standard E)	1	1		3	3		1	7	10
The Earth (Standard F)	1			1	4	1	1	10	11
The Universe (Standard G)	2	1	1	8	7	1		9	17
Energy (Standard H)	1			1	6	2		10	11
Motion (Standard I)	2		1	6	4		1	8	14
Application	8	2	2	20	28	5	5	58	78
Inquiry and Problem Solving (Standard J)	2	1	1	8	7	1	1	13	21
Scientific Reasoning (Standard K)	2	1		4	7	3		13	17
Communication (Standard L)	3		1	7	9	1	2	19	26
Implications of Science and Technology (Standard M)	1			1	5		2	13	14

SCIENCE AND TECHNOLOGY
NUMBER OF POINTS POSSIBLE
GRADE 8

Standard	Common					Matrix				Total Possible Points
	MC X1	SA X2	CR X4	ER x8	Points	MC X1	SA X2	CR X4	Points	
Content	18	4		.5	30	49	7	5	83	113
Classifying Life Forms (Standard A)						7	1		9	9
Ecology (Standard B)	2				2	5	1		7	9
Cells (Standard C)	4				4	5	2	1	13	17
Continuity and Change (Standard D)	2	1			4	8		1	12	16
Structure of Matter (Standard E)	3	1			5	4	1	1	10	15
The Earth (Standard F)				.5	4	5	1	1	11	15
The Universe (Standard G)	2	1			4	5		1	9	13
Energy (Standard H)	3	1			5	3			3	8
Motion (Standard I)	2				2	7	1		9	11
Application	2	1	3	.5	20	23	5	7	61	81
Inquiry and Problem Solving (Standard J)	1		1		5	7		3	19	24
Scientific Reasoning (Standard K)	1	1	1		7	5	3	2	19	26
Communication (Standard L)			1		4	6	1		8	12
Implications of Science and Technology (Standard M)				.5	4	5	1	2	15	19

SCIENCE AND TECHNOLOGY
NUMBER OF POINTS POSSIBLE
GRADE 11

Standard	Common					Matrix				Total Possible Points
	MC X 1	SA X 2	CR X 4	ER x 8	Points	MC X 1	SA X 2	CR X 4	Points	
Content	12	5	1	.5	30	51	7	5	85	115
Classifying Life Forms (Standard A)		1			2	6			6	8
Ecology (Standard B)	1				1	6	1	1	12	13
Cells (Standard C)	3	1			5	4	1	1	10	15
Continuity and Change (Standard D)	2				2	5	1		7	9
Structure of Matter (Standard E)	2	1			4	7		1	11	15
The Earth (Standard F)				.5	4	7	2	1	15	19
The Universe (Standard G)	1	1			3	4	1	1	10	13
Energy (Standard H)	2	1			4	5			5	9
Motion (Standard I)	1		1		5	7	1		9	14
Application	8		2	.5	20	21	5	7	59	79
Inquiry and Problem Solving (Standard J)	3			.5	7	10		1	14	21
Scientific Reasoning (Standard K)			1		4	2	3	1	12	16
Communication (Standard L)	5				5	9	2	2	21	26
Implications of Science and Technology (Standard M)			1		4			3	12	16

CHAPTER 7—DESIGN OF THE SOCIAL STUDIES ASSESSMENT

BLUEPRINT

The social studies framework was based on Maine's *Learning Results*, which identifies a total of thirteen **content standards** in the four disciplines—civics and government, history, geography, and economics—as listed below:

CIVICS AND GOVERNMENT

- **Rights, responsibilities, and participation:** Students understand the rights and responsibilities of civic life and employ the skills of effective civic participation.

- **Purpose and types of government:** Students understand the types and purposes of governments, their evolution, and their relationships with the governed.

- **Fundamental principles of government and constitutions:** Students understand the constitutional principles and the democratic foundations of the political institutions of the United States.

- **International relations:** Students understand the political relationships among the United States and other nations.

HISTORY

- **Chronology:** Students use the chronology of history and major eras to demonstrate the relationships of events and people.

- **Historical knowledge, concepts, and patterns:** Students develop historical knowledge of major events, people, and enduring themes in the United States, in Maine, and throughout world history.

- **Historical inquiry, analysis, and interpretation:** Students learn to evaluate resource material such as documents, artifacts, maps, artwork, and literature, and to make judgments about the perspectives of the authors and their credibility when interpreting current historical events.

- **Skills and tools:** Students know how to construct and interpret maps and use globes and other geographic tools to locate and derive information about people, places, regions, and environments.

- **Human interaction with environments:** Students understand and analyze the relationships among people and their physical environments.

ECONOMICS

- **Personal and consumer economics:** Students understand that economic decisions are based on the availability of resources and the costs and benefits of choices.

- **Economic systems of the United States:** Students understand the economic system of the United States, including its principles, development, and institutions.

- **Comparative systems:** Students analyze how different economic systems function and change over time.

- **International trade and global interdependence:** Students understand the patterns and results of international trade.

These thirteen standards have been used to create the reporting category framework for social studies, shown below.

Social Studies Framework		
Standard	**Percentage of Questions Emphasizing Content**	**Percentage of Questions Emphasizing Application**
Civics and Government:		
A. Rights, Responsibilities, and Participation	50%	50%
B./C. Purposes, Types, and Fundamental Principles	60%	40%
D. International Relations	60%	40%
History:		
A./B. Chronology and Historical Knowledge, Concepts, and Patterns	60%	40%
C. Historical Inquiry, Analysis, and Interpretation	40%	60%
Geography:		
A. Skills and Tools	40%	60%
B. Human Interaction with Environments	60%	40%
Economics:		
A. Personal and Consumer Economics	50%	50%
B./C. Economic Systems	50%	50%
D. International Trade and Global Interdependence (Grades 8 and 11)	60%	40%

Social studies education stresses a strong commitment to content knowledge, emphasizes the student's ability to engage in complex thinking and reasoning skills, and emphasizes the clear communication of ideas. Social studies assessment focuses on both content and applications to evaluate what students know and can demonstrate.

CONTENT SPECIFICATIONS

The MEA social studies assessment included multiple-choice, short-answer, constructed-response, and extended-response questions. Short-answer questions, which were new in the revised MEA, required students to answer questions using one or two words or a short phrase. Extended-response questions in social studies are similar to constructed-response questions except that they are more complex, requiring 10-20 minutes of response time. Each type of question was worth a specific number of points in the student's total social studies score, as shown below.

Type of Question	Possible Score Points
Multiple Choice	0–1
Short Answer	0–2
Constructed Response	0–4
Extended Response	0–8 (grades 8 and 11)

TEST DESIGN

The tables below summarize the numbers and types of questions that were used in the 2000-01 social studies assessment.

GRADE 4

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
3A	7	1	2	0	2	1	0	0	30 (+10)
3B	7	2	1	0	2	0	1	0	30 (+10)
3C	6	2	2	0	2	0	0	0	30 (+10)

GRADES 8/11

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
3A	13	1	1	1	2	1	0	0	50 (+15)
3B	7	4	2	0	4	0	1	0	50 (+15)

Key

- MC = multiple-choice questions
- SA = short-answer questions
- CR = constructed-response questions
- ER = extended-response questions

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

SOCIAL STUDIES
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common				Matrix				Total Possible Points 194
	MC x 1	SA x 2	CR x 4	Points	MC x 1	SA x 2	CR x 4	Points	
Content	17	2	1	25	68	3	1	78	103
Application	3	3	4	25	4	9	11	66	91
Civics and Government (Standards A, B, and C)	9	1	2	19	19	3	3	37	56
Rights, Responsibilities, and Participation (Standard A)	4		1	8	10	2	1	18	26
Purpose, Types, and Fundamental Principles of Government (Standards B and C)	5	1	1	11	9	1	2	19	30
History (Standards A, B, and C)	3	1	2	13	23	4	3	43	56
Chronology, Historical Knowledge, Concepts, and Patterns (Standards A and B)	3	1	1	9	21	2	1	29	38
Historical Inquiry, Analysis, and Interpretation (Standard C)			1	4	2	2	2	14	18
Geography (Standards A and B)	5	2	1	13	22	3	4	44	57
Skills and Tools (Standard A)	5	1	1	11	10	1	2	20	31
Human Interaction with Environments (Standard B)		1		2	12	2	2	24	26
Economics (Standards A and B)	3	1		5	8	2	2	20	25
Personal and Consumer Economics/ Economic Systems (Standards A and B)	3	1		5	8	2	2	20	25

SOCIAL STUDIES
NUMBER OF POINTS POSSIBLE
GRADE 8

Standard	Common					Matrix				Total Possible Points 193
	MC x1	SA x2	CR x4	ER x8	Points	MC x1	SA x2	CR x4	Points	
Content	19	4			27	71	10		91	118
Application		1	3	1	22	1	2	12	53	75
Civics and Government (Standards A, B, C, and D)	5	2	1		13	18	3	3	36	49
Rights, Responsibilities, and Participation (Standard A)	2		1		6	6		1	10	16
Purpose, Types, and Fundamental Principles of Government (Standards B and C)	2	2			6	8	2	1	16	22
International Relations (Standard D)	1				1	4	1	1	10	11
History (Standards A, B, and C)	5	1		1	15	21	3	4	43	58
Chronology, Historical Knowledge, Concepts, and Patterns (Standards A and B)	5	1		1	15	20	2	2	32	47
Historical Inquiry, Analysis, and Interpretation (Standard C)						1	1	2	11	11
Geography (Standards A and B)	5	1	1		11	18	3	3	36	47
Skills and Tools (Standard A)	3				3	7	2	1	15	18
Human Interaction with Environments (Standard B)	2	1	1		8	11	1	2	21	29
Economics (Standards A, B, and D)	4	1	1		10	15	3	2	29	39
Personal and Consumer Economics (Standard A)		1			2	6			6	8
Economic Systems/Comparative Systems (Standards B and C)	2		1		6	9	1	1	15	21
International Trade and Global Interdependence (Standard D)	2				2		2	1	8	10

SOCIAL STUDIES
NUMBER OF POINTS POSSIBLE
GRADE 11

Standard	Common					Matrix				Total Possible Points 194
	MC x 1	SA x 2	CR x 4	ER x 8	Points	MC x 1	SA x 2	CR x 4	Points	
Content	14	4			22	54	9		72	94
Application	6	1	3	1	28	18	3	12	72	100
Civics and Government (Standards A, B, C, and D)	5	1	1		11	16	4	3	36	47
Rights, Responsibilities, and Participation (Standard A)	2				2	2	1	2	12	14
Purpose, Types, and Fundamental Principles of Government (Standards B and C)	3		1		7	10	2	1	18	25
International Relations (Standard D)		1			2	4	1		6	8
History (Standards A, B, and C)	5	2		1	17	23	2	4	43	60
Chronology, Historical Knowledge, Concepts, and Patterns (Standards A and B)	5	1			7	21	2	2	33	40
Historical Inquiry, Analysis, and Interpretation (Standard C)		1		1	10	2		2	10	20
Geography (Standards A and B)	5	1	1		11	18	3	3	36	47
Skills and Tools (Standard A)	3				3	10	1	2	20	23
Human Interaction with Environments (Standard B)	2	1	1		8	8	2	1	16	24
Economics (Standards A, B, and D)	5	1	1		11	15	3	2	29	40
Personal and Consumer Economics (Standards A)	1				1	2	1	1	8	9
Economic Systems/Comparative Systems (Standards B and C)	3	1	1		9	10	1		12	21
International Trade and Global Interdependence (Standard D)	1				1	3	1	1	9	10

CHAPTER 8—DESIGN OF THE VISUAL AND PERFORMING ARTS ASSESSMENT

BLUEPRINT

The visual and performing arts assessment includes four disciplines: dance, music, theater, and visual arts. The arts framework is based on Maine's *Learning Results*, which identifies three content standards in the arts as listed below:

- **Creative expression:** Students create and/or perform to express ideas and feelings.

- **Cultural heritage:** Students understand the cultural contributions (social, ethical, political, religious dimensions) of the arts, how the arts shape and are shaped by prevailing cultural and social beliefs and values, and recognize exemplary works from a variety of cultures and historical periods.

- **Criticism and aesthetics:** Students reflect upon and assess the characteristics and merits of art works.

These three standards were used to create the reporting category framework for the visual and performing arts, as shown below.

Visual and Performing Arts Framework

Discipline	Standard		
	A. Creative Expression	B. Cultural Heritage	C. Criticism and Aesthetics
Dance			
Music			
Theater			
Visual Arts			

Each row and each column of the framework constitutes a reporting category for school- and district-level results in the MEA—for example, music/cultural heritage. Student-level results were not reported in the visual and performing arts, as no common items were used in this area.

It should be noted that not all of the performance indicators associated with each content standard (see *Learning Results*) can be assessed reliably and validly using a paper-and-pencil test. For example, some of the

performance indicators included under the standard for "creative expression" would best be measured in other ways. For this reason, additional methods of assessment for these performance indicators are being studied.

The distribution of questions, or emphasis, across the arts disciplines in the MEA varies from one grade level to another, as shown in the table below.

Discipline	Grade		
	4	8	11
Dance	10%	10%	15%
Music	40%	40%	35%
Theater	10%	10%	15%
Visual Arts	40%	40%	35%

CONTENT SPECIFICATIONS

The MEA visual and performing arts assessment included multiple-choice and constructed-response questions. Each type of question was worth a specific number of points, as shown below:

Type of Question	Possible Score Points
Multiple Choice	0–1
Constructed Response	0–4

TEST DESIGN

The table below summarizes the numbers and types of matrix-sampled questions that were used in the 2000-01 visual and performing arts assessment.

Visual and Performing Arts

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
5A					6	0	1	0	20 (+10)

Key

- MC = multiple-choice questions
- CR = constructed-response questions

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

VISUAL AND PERFORMING ARTS
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common			Matrix			Total Possible Points 119
	MC	CR	Points	MC x 1	CR x 4	Points	
Dance				15	3	27	27
Music				24	3	36	36
Theater				14	3	26	26
Visual Arts				18	3	30	30
Creative Expression (Standard A)				31	4	47	47
Cultural Heritage (Standard B)				18	3	30	30
Criticism and Aesthetics (Standard C)				22	5	42	42

VISUAL AND PERFORMING ARTS
NUMBER OF POINTS POSSIBLE
GRADE 8

Standard	Common			Matrix			Total Possible Points 117
	MC	CR	Points	MC x 1	CR x 4	Points	
Dance				10	3	22	22
Music				24	3	36	36
Theater				11	3	23	23
Visual Arts				24	3	36	36
Creative Expression (Standard A)				30	4	46	46
Cultural Heritage (Standard B)				16	4	32	32
Criticism and Aesthetics (Standard C)				23	4	39	39

VISUAL AND PERFORMING ARTS
NUMBER OF POINTS POSSIBLE
GRADE 11

Standard	Common			Matrix			Total Possible Points 118
	MC	CR	Points	MC x 1	CR x 4	Points	
Dance				12	3	24	24
Music				24	3	36	36
Theater				12	3	24	24
Visual Arts				22	3	34	34
Creative Expression (Standard A)				28	5	48	48
Cultural Heritage (Standard B)				20	3	32	32
Criticism and Aesthetics (Standard C)				22	4	38	38

CHAPTER 9—DESIGN OF THE HEALTH EDUCATION ASSESSMENT

BLUEPRINT

The health framework was based on Maine's *Learning Results*, which identifies six **content standards** as shown below:

- **Health concepts:** Students understand health promotion and disease prevention concepts.

- **Health information, services, and products:** Students know how to acquire valid information about health issues, services, and products.

- **Health promotion and risk reduction:** Students understand how to reduce their health risks through the practice of healthy behaviors.

- **Influences on health:** Students understand how media techniques, cultural perspectives, technology, peers, and family influence behaviors that affect health.

- **Communication skills:** Students understand that skillful communication can contribute to better health for them, their families, and the community.

- **Decision making and goal setting:** Students learn how to set personal goals and make decisions that lead to better health.

These six standards were combined with the ten health education content areas identified by the 1984 Education Reform Act to create a reporting category framework for health, as shown on the next page.

Health Framework						
	Health Standard					
Content Area	A. Health Concepts	B. Health Information, Services, and Products	C. Health Promotion and Risk Reduction	D. Influences on Health	E. Communication Skills	F. Decision Making and Goal Setting
Community, Consumer, and Environmental Health						
Personal and Nutritional Health						
Family Life Education and Growth and Development						
Safety and Injury Prevention						
Tobacco, Alcohol, and Other Drug Use Prevention						
Prevention and Control of Disease and Disorders						
Total	30%	70%				

Thirty percent of the questions measured health standard A; they were divided among the six content areas. The remaining 70% of the questions were divided among standards B through F and the six content areas. The distribution of questions was 10% to 20% for each standard, determined by its developmental appropriateness for the specific grade being assessed.

A portion of the questions in the health assessment were developed by the Health Education Assessment Project for the State Collaborative on Assessment and Student Standards (SCASS) under the auspices of the Council of Chief State School Officers. Each SCASS question that was used or adapted was aligned with a performance indicator from Maine's health education standards. Maine educators on the content development committee developed the remainder of the questions.

CONTENT SPECIFICATIONS

The MEA health assessment included multiple-choice, short-answer, constructed-response, and extended-response questions (grades 8 and 11 only). Short-answer questions, which were new in the revised MEA, required students to formulate answers using one or two words or a short phrase. Extended-response questions in health are similar to constructed-response questions except that they are more complex, requiring 10–20 minutes of response time. Each type of question was worth a specific number of points in the student's total health score, as shown below.

Type of Question	Possible Score Points
Multiple Choice	0–1
Short Answer	0–2
Constructed Response	0–4
Extended Response	0–8 (grades 8 and 11)

TEST DESIGN

At every grade level, the assessment included no common questions but was constructed solely of matrix-sampled questions. The tables below summarize the numbers and types of questions that were used in the 2000-01 health education assessment for each grade.

GRADE 4

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
4A					6	1	3	0	40 (+15)

GRADES 8/11

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
4A					6	1	1	1	40 (+15)

Key

- MC = multiple-choice questions
- SA = short-answer questions
- CR = constructed-response questions
- ER = extended-response questions

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

HEALTH EDUCATION
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common			Matrix			Total Possible Points 236
	MC	SA	CR	MC X1	SA X2	CR X4	
Health Concepts (Standard A)				22	5	9	68
Health Information, Services, and Products (Standard B)				15	2	3	31
Health Promotion and Risk Reduction (Standard C)				16	3	7	50
Influences on Health (Standard D)				11	2	3	27
Communication Skills (Standard E)				2	0	8	34
Decision Making and Goal Setting (Standard F)				6	0	5	26
Community, Consumer, and Environmental Health				12	2	6	40
Personal and Nutritional Health				18	3	9	60
Family Life Education and Growth and Development				8	1	7	38
Safety and Injury Prevention				13	2	7	45
Tobacco, Alcohol, and Other Drug Use Prevention				11	1	4	29
Prevention and Control of Disease and Disorders				10	3	2	24

HEALTH EDUCATION
NUMBER OF POINTS POSSIBLE
GRADE 8

Standard	Common				Matrix				Total Possible Points 240
	MC	SA	CR	ER	MC X 1	SA X 2	CR X 4	ER X 8	
Health Concepts (Standard A)					40	6	1	3	80
Health Information, Services, and Products (Standard B)					7	1	2	1	25
Health Promotion and Risk Reduction (Standard C)					15	2	2	2	43
Influences on Health (Standard D)					3	1	4	1	29
Communication Skills (Standard E)					4	2	1	3	36
Decision Making and Goal Setting (Standard F)					3	0	2	2	27
Community, Consumer, and Environmental Health					10	2	4	0	30
Personal and Nutritional Health					8	2	0	4	44
Family Life Education and Growth and Development					13	4	1	3	49
Safety and Injury Prevention					18	1	3	2	48
Tobacco, Alcohol, and Other Drug Use Prevention					15	1	3	2	45
Prevention and Control of Disease and Disorders					8	2	1	1	24

HEALTH EDUCATION
NUMBER OF POINTS POSSIBLE
GRADE 11

Standard	Common				Matrix				Total Possible Points 240
	MC	SA	CR	ER	MC X 1	SA X 2	CR X 4	ER X 8	
Health Concepts (Standard A)					28	5	3	5.5	94
Health Information, Services, and Products (Standard B)					9	3	0	2	31
Health Promotion and Risk Reduction (Standard C)					13	1	3	2	43
Influences on Health (Standard D)					10	1	1	1	24
Communication Skills (Standard E)					5	2	2	1	25
Decision Making and Goal Setting (Standard F)					7	0	3	.5	23
Community, Consumer, and Environmental Health					8	1	2	2	34
Personal and Nutritional Health					19	1	3	1	41
Family Life Education and Growth and Development					10	2	3	3	50
Safety and Injury Prevention					14	3	2	1	36
Tobacco, Alcohol, and Other Drug Use Prevention					10	2	2	3	46
Prevention and Control of Disease and Disorders					11	3	0	2	33

SECTION II: TEST ADMINISTRATION
CHAPTER 10—TEST ADMINISTRATION

RESPONSIBILITY FOR ADMINISTRATION

As indicated in the *Principal/Test Coordinator's Manual*, principals and/or their designated MEA coordinator were responsible for the proper administration of the MEA. Manuals and certification forms were used to ensure the uniformity of administration procedures from school to school.

PROCEDURES

Principals and/or the school's designated MEA coordinator were instructed to read the *Principal/Test Coordinator's Manual* prior to testing and to be familiar with the instructions given in the *Test Administrator's Manual*. The *Principal/Test Coordinator's Manual* provided each school with checklists to help them to prepare for testing. The checklists outlined tasks for the schools to perform before, during, and after test administration. Along with these checklists, the *Principal/Test Coordinator's Manual* outlined the nature of the testing material being sent to each school, how to inventory the material, how to track it during administration, and how to return the material once testing was complete. It also contained information about including or excluding students. The *Test Administrator's Manual* also included checklists for the administrators to prepare themselves, their classrooms, and the students for the administration of the test. The *Test Administrator's Manual* contained sections that detailed the procedures to be followed for each test session, and it contained instructions on preparing the material prior to giving it to the principal/coordinator for its return to Measured Progress.

ADMINISTRATOR TRAINING

In addition to distributing the *Principal/Test Coordinator's* and *Test Administrator's Manual*s, the Maine Department of Education, along with Measured Progress, conducted two ITV workshops (one in the fall and one in the winter) to train and inform school personnel about the revised MEA.

PARTICIPATION REQUIREMENTS

The following categories of students were allowed to be considered for modifications:

- Students who had an identified exceptionality/disability

- Students who had been identified as limited English proficient (LEP)

- Students who were unable to work independently in any of the subjects assessed

- Students who were ill or incapacitated in some way

All students who were considered for modifications on the MEA were to have had their individual situations reviewed by a group within the school prior to the time of testing. For every student with an identified exceptionality requiring an Individual Educational Plan (IEP), schools were required to hold a Pupil Evaluation Team (PET) meeting that addressed that student's needs for modifications. Other students needing test modifications, who did not have an identified exceptionality, were required to attend a meeting that included one of the student's teachers, the building principal, related services personnel, and, whenever possible, the student's parents. If it was not possible for the parents to attend the meeting, it was required that they be notified of the committee's recommendations for modifications prior to the time of testing.

Recommended modifications were to be consistent with those modifications already being employed in the student's instructional program. Any such modifications were reflected either in the minutes of the PET meeting (for students requiring an IEP) or in a statement prepared for the cumulative folders of students not requiring IEPs. The following is the suggested statement that schools were given as a model:

The student will/will not participate in the __th-grade Maine Educational Assessment as scheduled during the month of _____ 19__. The following test modifications will be observed: (list modifications)

EXCLUSION FROM THE ASSESSMENT

Exclusion was defined as the most extreme modification of the assessment. Since it was clear that the legislation's intent was to include as many students as possible, it was recommended that exclusion be considered only as a last resort.

On those occasions where it was deemed necessary to exclude a student from sections of the assessment or from the assessment as a whole, it was recommended that exclusion be limited to only those sections of the MEA that were considered inappropriate for that particular student. Exclusion was to be selected only after the various types of modifications available had been fully explored, and it was felt that the assessment would not yield a valid indication of how a student functioned in a given content area. For example, even students who were reading two years below grade level were advised to take the reading section because those scores would give a fair representation of their current level of functioning in reading. If, however, after examining all of the possible modifications, a local school decided that the assessment or sections of it would be inappropriate for a given student, that student could be excluded.

STUDENTS ENROLLED IN UNGRADED OR MULTI-AGE PROGRAMS

For the purposes of the assessment, it was recommended that students enrolled in ungraded or multi-age programs be tested with the fourth grade if they were 9 years old, with the eighth grade if they were 13, and with the eleventh grade if they were 17.

DOCUMENTATION OF MODIFICATIONS OR EXCLUSIONS

Information about the modifications given to students or the reasons for exclusion was to be provided on the front page of the student's response booklet. This information was to be coded in by staff, not students, after testing was completed. The *Test Coordinator's* and *Test Administrator's Manual* provided directions on coding in the information related to modification(s), partial exclusion, and exclusion, and every student who was totally excluded had to be accounted for in the designated section of the response booklet.

STATE PARTICIPATION RATES—FALL 2000

GRADE 4

Student Participation Category	Number	Percentage
Students Enrolled: number of completed test booklets	16321	100
Total Students Not Included in Report(s):	1336	8
students who took no session of the assessment due to an identified disability	392	2
students who took some but not all sessions of the assessment due to an identified disability	362	2
students tested who receive special education and related services for more than 60% of the school day in a self-contained classroom as defined in Maine Special Education Regulations, Chapter 101, 11.6	117	1
students who took no session of the assessment due to LEP	41	0
students who took some but not all sessions of the assessment due to LEP	2	0
students who took no session of the assessment due to 504 Plan, absence, or other approved reason	63	0
students who took some but not all sessions of the assessment due to 504 Plan, absence, or other approved reason	359	2
Total Students Completing All Subjects:	14985	92
students with identified disability completing all subjects without accommodations	253	2
students with identified disability completing all subjects with accommodations	1125	7
all others completing all subjects	13607	83
Percentage of Students with Identified Disability Included in Reports for All Subjects:		9
Percentage of All Other Students Included in Reports for All Subjects:		91

GRADE 8

Student Participation Category	Number	Percentage
Students Enrolled: number of completed test booklets	17162	100
Total Students Not Included in Report(s):	1217	7
students who took no session of the assessment due to an identified disability	379	2
students who took some but not all sessions of the assessment due to an identified disability	100	1
students tested who receive special education and related services for more than 60% of the school day in a self-contained classroom as defined in Maine Special Education Regulations, Chapter 101, 11.6	211	1
students who took no session of the assessment due to LEP	22	0
students who took some but not all sessions of the assessment due to LEP	1	0
students who took no session of the assessment due to 504 Plan, absence, or other approved reason	221	1
students who took some but not all sessions of the assessment due to 504 Plan, absence, or other approved reason	283	2
Total Students Completing All Subjects:	15945	93
students with identified disability completing all subjects without accommodations	326	2
students with identified disability completing all subjects with accommodations	1218	7
all others completing all subjects	14401	84
Percentage of Students with Identified Disability Included in Reports for All Subjects:		10
Percentage of All Other Students Included in Reports for All Subjects:		90

GRADE 11

Student Participation Category	Number	Percentage
Students Enrolled: number of completed test booklets	15290	100
Total Students Not Included in Report(s):	1363	9
students who took no session of the assessment due to an identified disability	277	2
students who took some but not all sessions of the assessment due to an identified disability	38	0
students tested who receive special education and related services for more than 60% of the school day in a self-contained classroom as defined in Maine Special Education Regulations, Chapter 101, 11.6	90	1
students who took no session of the assessment due to LEP	34	0
students who took some but not all sessions of the assessment due to LEP	2	0
students who took no session of the assessment due to 504 Plan, absence, or other approved reason	446	3
students who took some but not all sessions of the assessment due to 504 Plan, absence, or other approved reason	476	3
Total Students Completing All Subjects:	13927	91
students with identified disability completing all subjects without accommodations	246	2
students with identified disability completing all subjects with accommodations	691	5
all others completing all subjects	12990	85
Percentage of Students with Identified Disability Included in Reports for All Subjects:		7
Percentage of All Other Students Included in Reports for All Subjects:		93

STATE PARTICIPATION RATES—SPRING 2001

GRADE 4

Student Participation Category	Number	Percentage
Students Enrolled: number of completed test booklets	16328	100
Total Students Not Included in Report(s):	778	5
students who took no session of the assessment due to an identified disability	378	2
students who took some but not all sessions of the assessment due to an identified disability	63	0
students tested who receive special education and related services for more than 60% of the school day in a self-contained classroom as defined in Maine Special Education Regulations, Chapter 101, 11.6	105	1
students who took no session of the assessment due to LEP	0	0
students who took some but not all sessions of the assessment due to LEP	3	0
students who took no session of the assessment due to 504 Plan, absence, or other approved reason	55	0
students who took some but not all sessions of the assessment due to 504 Plan, absence, or other approved reason	174	1
Total Students Completing All Subjects:	15550	95
students with identified disability completing all subjects without accommodations	231	1
students with identified disability completing all subjects with accommodations	1626	10
all others completing all subjects	13693	84
Percentage of Students with Identified Disability Included in Reports for All Subjects:		12
Percentage of All Other Students Included in Reports for All Subjects:		88

GRADE 8

Student Participation Category	Number	Percentage
Students Enrolled: number of completed test booklets	17093	100
Total Students Not Included in Report(s):	1113	7
students who took no session of the assessment due to an identified disability	358	2
students who took some but not all sessions of the assessment due to an identified disability	48	0
students tested who receive special education and related services for more than 60% of the school day in a self-contained classroom as defined in Maine Special Education Regulations, Chapter 101, 11.6	218	1
students who took no session of the assessment due to LEP	0	0
students who took some but not all sessions of the assessment due to LEP	3	0
students who took no session of the assessment due to 504 Plan, absence, or other approved reason	186	1
students who took some but not all sessions of the assessment due to 504 Plan, absence, or other approved reason	300	2
Total Students Completing All Subjects:	15980	93
students with identified disability completing all subjects without accommodations	338	2
students with identified disability completing all subjects with accommodations	1311	8
all others completing all subjects	14331	84
Percentage of Students with Identified Disability Included in Reports for All Subjects:		10
Percentage of All Other Students Included in Reports for All Subjects:		90

GRADE 11

Student Participation Category	Number	Percentage
Students Enrolled: number of completed test booklets	14946	100
Total Students Not Included in Report(s):	1341	9
students who took no session of the assessment due to an identified disability	271	2
students who took some but not all sessions of the assessment due to an identified disability	28	0
students tested who receive special education and related services for more than 60% of the school day in a self-contained classroom as defined in Maine Special Education Regulations, Chapter 101, 11.6	97	1
students who took no session of the assessment due to LEP	2	0
students who took some but not all sessions of the assessment due to LEP	0	0
students who took no session of the assessment due to 504 Plan, absence, or other approved reason	421	3
students who took some but not all sessions of the assessment due to 504 Plan, absence, or other approved reason	522	3
Total Students Completing All Subjects:	13605	91
students with identified disability completing all subjects without accommodations	243	2
students with identified disability completing all subjects with accommodations	701	5
all others completing all subjects	12661	85
Percentage of Students with Identified Disability Included in Reports for All Subjects:		7
Percentage of All Other Students Included in Reports for All Subjects:		93

TESTING IRREGULARITIES

There were **no** testing irregularities for the 2000–01 assessment year.

SECTION III: DEVELOPMENT AND REPORTING OF SCORES

CHAPTER 11—SCORING

MACHINE SCORED ITEMS

Once the 2000-01 booklets had been logged in, identified with appropriate scannable, pre-printed school information sheets, examined for extraneous materials, and batched, they were moved into the scanning area. For all response booklets (and questionnaires and other forms that require imaging/scanning) to be imaged, this area is the last stop in the processing loop in which the documents themselves are handled.

At that point, 100% of the response document and other scannable information necessary to produce the required reports had been captured and converted into an electronic format, including all student identification and demographics, selected-response answers, and digital image clips of hand-written responses. The digital image clip information allowed Measured Progress to replicate student responses just as they appeared on the originals, but they had been transferred onto the readers' monitors. From that point on, the entire process—data processing, scoring, "range-finding," data analysis, reporting—was accomplished without further reference to the originals.

The first step in that conversion was the removal of the booklet bindings so that the individual pages could pass through the scanners, one at a time. Once cut, the sheets were put back in their proper boxes and placed in storage until needed for the scanning/imaging process.

Customized scanning programs for all scannables were prepared to selectively read the student response booklets and to format the scanned information electronically according to pre-determined requirements. Any information (including multiple-choice response data) that had been designated time-critical or process-critical was handled first.

In addition to numerous real-time quality control checks, duplex read, a transport printer that prints a unique identifying number on each sheet of each booklet, and on-line editing capability, the 5000i scanners offer features that make them compatible with Internet technology.

SCANNING QUALITY CONTROL

NCS scanners are equipped with many built-in safeguards that prevent data errors. The scanning hardware is continually monitored for conditions that will cause the machine to shut down if standards are not met. It will display an error message and prevent further scanning until the condition is corrected. The areas monitored include document page and integrity checks, user-designed on-line edits, and many internal checks of electronic functions.

Before every scanning shift begins, Measured Progress's operators performed a daily diagnostic routine. This is yet another step to protect data integrity, and one that has been done faithfully for the many years that we have been involved in production scanning. In the rare event that the routine detects a photocell that appears to be out of range, we calibrate that machine and perform the test again. If the read is still not up to standard, we call for assistance from our field service engineer.

As a final safeguard, spot checks of scanned files, bubble by bubble and image by image, were routinely made throughout scanning runs. The result of these precautions, from the original layout of the scanning form to the daily vigilance of our operators, was a scan error rate well below 0.001.

ELECTRONIC DATA FILES

Once the data had been entered and the scanning logs and other paperwork completed, the booklets themselves were put into storage (where they stayed for at least 180 days beyond the close of the fiscal year). When it had been determined that the files were complete and accurate, those files were duplicated electronically and made available for many other processing options. Completed files were loaded onto our local area network (LAN) for transfer to Measured Progress' proprietary I-Score system for scoring. Those files were then used to identify (and print out) papers to be used in the rangefinding and standard-setting processes and the data was made transferable via the Internet, CD-ROM, or optical disk.

ITEMS SCORED BY READERS

Test and answer materials were handled as little as possible to minimize the possibility of loss, mishandling, or breach of security. Once scanned, either by optical mark reader or the I-Score system, papers were stored securely in areas with limited personnel access.

As explained in the following sections on scoring, the I-Score system itself ensures the security of responses and test items: all scoring is "blind"; that is, no student names are associated with viewed responses or raw scores and all scoring personnel are subject to the same nondisclosure requirements and supervision as regular Measured Progress staff.

I-SCORE

After the 2000-01 test material had been loaded into the LAN, I-Score sent electronically scanned images of student work to individual readers at computer terminals who evaluated each response and recorded each student's score via keypad or mouse entry. When the reader had finished with one response, the next response appeared immediately on the computer screen. In that way, the system guaranteed complete anonymity of individual students and ensured the randomization of responses during scoring.

Although I-Score is based on conventional scoring techniques, it also offers numerous benefits, not the least of which is raising the bar on scoring process capability. Some of the benefits are as follows:

- real-time information on scorer reliability, read-behinds, and overall process monitoring;

- early access to subsets of data for tasks such as standard setting;

- reduced material handling, which not only saves time and labor, but also enhances the security of materials; and

- immediate access to samples of student responses and scores for reporting and analysis through electronic media.

Scoring operations, directed by the manager of scoring services, are carried out by a highly qualified staff. The staff included:

- chief readers, who oversaw all training and scoring within particular subject areas;

- quality assurance coordinators (QACs), who lead rangefinding and training activities and monitor scoring consistency and rates;

- verifiers, who perform read-behinds of readers and assist at scoring tables as necessary; and

- readers, who perform the bulk of the scoring.

Table 11-1 summarizes the qualifications of the 2000-01 MEA quality assurance coordinators and readers.

Table 11-1 Qualifications of 2000-01 QACs and Readers					
2000 Fall Administration					
Scoring Responsibility	Educational Credentials				Total
	Doctorate	Masters	Bachelors	Other	
QACs		100.0			100%
Readers	2.8	36.1	44.4	16.7	100%
2001 Spring Administration					
Scoring Responsibility	Educational Credentials				Total
	Doctorate	Masters	Bachelors	Other	
QACs	13.3	53.3	20.0	13.4	100%
Readers	2.4	19.0	48.2	30.4	100%

PRELIMINARY ACTIVITIES

Preliminary activities for scoring included (1) participating in the planning and design of documents to be used for scoring, (2) reviewing items and score guides for rangefinding and training and the creation of rangefinding packets, and (3) selecting scoring staff and training them for scoring.

PLANNING AND DESIGNING DOCUMENTS

At the request of Measured Progress' project manager, scoring personnel advised project management and DOE staff on the program design in order to support an efficient and effective scoring process. Scoring staff contributed also to the design of

- response documents and the image-capture process to yield acceptable image clips (also defining file format and layout); and

- scoring benchmarks composed of the guide, subject background information, and anchor papers.

REVIEWING ITEMS AND GUIDES (RANGEFINDING)

Before the scheduled start of scoring activities, scoring center staff reviewed test items and scoring guides for rangefinding. At that point, chief readers and selected QACs prepared scorer training materials. Measured Progress's scoring staff (including test developers) selected one or two anchor examples for each item score point. An additional six to ten responses per item were chosen as part of the training pack. The anchor pack consisted of mid-range exemplars, while the training pack exemplars illustrated the range within each score point. The chief readers, who worked closely with QACs for each content area, facilitated the selection of response exemplars. One

of the greatest difficulties in the selection of anchor and training exemplars was finding a sufficient number of papers representing the highest scores (4 or 8) as such scores are fairly rare.

SELECTING AND TRAINING SCORING STAFF

SELECTING QUALITY ASSURANCE COORDINATORS (QACS) AND VERIFIERS

Because the read-behinds performed by the QACs and verifiers moderated the scoring process and thus maintained the integrity of the scores, individuals to fill those positions were selected for their accuracy. In addition, QACs, who train readers to score each item in their content areas, were selected for their ability to instruct and for their level of expertise in their content areas. For this reason, QACs typically are retired teachers who have demonstrated a high level of expertise in their respective disciplines. The ratio of QACs and verifiers to readers was approximately 1:11.

TRAINING QUALITY ASSURANCE COORDINATORS AND VERIFIERS

To ensure that all QACs provided consistent training and feedback, the chief readers spent two days training and qualifying the QACs, and the QACs reviewed all items with the verifiers before scoring. In addition, QACs rotated among tables, supervising readers and reading behind verifiers, who in turn read behind a different table of readers each day.

SELECTING READERS

Applicants were required to demonstrate their ability by participating in a preliminary scoring evaluation. The I-Score system enables Measured Progress to efficiently measure a prospective reader's ability to score student responses accurately. After having participated in a training session, applicants were required to achieve at least 80% exact scoring agreement for a qualifying pack consisting of 20 responses to a predetermined item in their content area. Those 20 responses were randomly selected from a bank of approximately 150, all of which had been selected by QACs and approved by the chief readers and developers.

TRAINING READERS

The QACs first applied the language of the scoring guide for an item to its anchor pack exemplars. Once discussion of the anchor pack had concluded, readers attempted to score the training pack exemplars correctly. The QACs then reviewed the training pack and answered any questions readers had before actual scoring began. With

this system, two aspects of scoring efficiency are in conflict. First, in order to minimize training expense, it is desirable to train each reader on as few items as possible. Second, to prevent reader drift and to minimize retraining requirements, it is desirable to score a given item in a brief period of time. But the lower the number of unique items each reader scores, the greater the number of readers required to score that item quickly. To minimize that conflict, we divided each subject area's readers into two or more groups. On the first day of scoring, each group was trained to score a different item. When a group had completed all of an item's responses, those readers were trained on another item (or set).

SCORING ACTIVITIES

Student test booklets at grade level 4 and student response booklets at grade levels 8 and 11 were digitally scanned and scored on a file server for a dedicated, secure LAN. I-Score then distributed digital images of student responses to readers. Training and scoring took place over a period of approximately two weeks. Items were randomly assigned to readers; thus, each item in a student's response booklet was more than likely scored by a different reader. By using the maximum possible number of readers for each student, the procedure effectively minimized error variance due to reader sampling. All common and matrix constructed- and extended-response items were scored once with a 2% read-behind to ensure consistency among readers and accuracy of individual readers.

MONITORING READERS

After a reader scored a student response, I-Score determined whether that response should also be scored by another reader, scored by a QAC or verifier, or routed for special attention. QACs and verifiers used I-Score to produce daily reader accuracy and speed reports. QACs and verifiers were able to obtain current reader accuracy reports and speed reports on-line at any time.

SCORING THE WRITING

Maine teachers and administrators were recruited to score the common writing prompt at in-state scoring sessions that were held in Bangor and Gorham, Maine. Teachers who participated in the scoring process developed skills in holistic evaluation of writing using a rubric aligned with the standards outlined in the Maine *Learning Results*. Those skills could then be applied to writing instruction in the classrooms, and the scoring of writing also gave participants an opportunity to read the range of student writing produced at each grade and to connect their current teaching practices

with the recommendations in the Maine *Learning Results*. Administrators who participated gained skills helpful in improving the teaching and evaluation of writing in their schools. Maine teachers' involvement in scoring also created a network of teachers who served as a resource to their local and state schools.

GENERAL SCORING GUIDES

SHORT-ANSWER ITEMS

Score Point	Description
2	▪ The student's response provides a complete and correct answer.
1	▪ The student's response is partially correct. ▪ The student's response may be incomplete or contain errors.
0	▪ The student's response is totally incorrect or too minimal to evaluate.
B	▪ Blank/no response.

CONSTRUCTED-RESPONSE ITEMS

Score Point	Description
4	▪ The student completes all important components of the task and communicates ideas clearly. ▪ The student demonstrates in-depth understanding of the relevant concepts and/or processes. ▪ When instructed to do so, the student chooses more efficient and/or sophisticated processes. ▪ When instructed to do so, the student offers insightful interpretations or extensions (e.g., generalizations, applications, and analogies).
3	▪ The student completes the most important components of the task and communicates clearly. ▪ The student demonstrates understanding of major concepts even though he/she overlooks or misunderstands some less important ideas or details.
2	▪ The student completes most important components of the task and communicates those clearly. ▪ The student demonstrates that there are gaps in his/her conceptual understanding.
1	▪ The student shows minimal understanding. ▪ The student addresses only a small portion of the required task(s).
0	▪ The student's response is totally incorrect or irrelevant.
B	▪ Blank/no response.

Stylistic & Rhetorical Aspects of Writing
Topic Idea Development

1	2	3	4	5	6
• Little topic development and/or organization, few details • Possible evidence of voice • Simplistic language (wording and sentence structures)	• Limited topic development, focus, and/or details • Evidence of voice • Limited variety in language used (wording and sentence structures)	• Moderate topic development, focus, details • Some voice • Some variety in language used (wording and sentence structures)	• Well developed with control and relevant details • Consistent voice • Variety in language used (wording and sentence structures)	• Fully developed with strong details • Sustained voice and/or tone with emerging style • Effective use of language	• Topic and details richly developed • Distinctive voice, tone and style • Rich use of language

Analytic Annotations

	Commendations	Needs
Topic Development The overall effect of the paper	TX sustained development throughout TY engages the reader	TJ more development of ideas/topic TK more direct response to the prompt
Organization The degree to which the response is: • Focused • Clearly and logically ordered • Clarified by paragraphs	OX clearly focused OY reflects careful planning	OJ a more focused middle OK a stronger ending
Details The degree to which the response includes examples that develop the main points	DX details support focus DY uses interesting details	DJ to avoid simply listing details DK more/relevant details
Language/Style The degree to which manipulation of language, including vocabulary, word choice, word combination, and sentence variety is effective	LX word choice enhances meaning LY sentence variety is used effectively	LJ to use more variety in language LK a clearer voice, tone, or style

Standard English Conventions

1	2	3	4
• Errors seriously interfere with communication and/or • Little control of sentence structure, grammar and usage, and mechanics in first draft writing	• Errors interfere somewhat with communication and/or • Few or no errors in simplistic or limited text in first draft writing	• Errors do not interfere with communication and/or • Few errors relative to length of essay or complexity of sentence structure, grammar and usage, and mechanics in first draft writing	• Control of a variety of sentence structures, grammar and usage, and mechanics • Length and complexity of essay provide opportunity for student to show control of standard English conventions in first draft writing

Analytic Annotations

	Commendations	Needs
Sentences The degree to which the response includes sentences that are correct in structure	SP correct sentence structure	SR correct sentence structure
Grammar and Usage The degree to which the response demonstrates correct • Use of standard grammatical rules of English • Word usage and vocabulary	GUP correct application of grammatical rules GUQ control of vocabulary and word usage	GUR correct application of grammatical rules GUS greater attention to correct word usage
Mechanics The degree to which the response demonstrates correct • Punctuation • Capitalization • Spelling	MP control of mechanics aids clarity MQ correct mechanics in sophisticated construction	MR greater control of mechanics e.g., spelling MS more careful editing, e.g., omissions, capitalization, and/or punctuation

CHAPTER 12—EQUATING AND SCALING

Scaled scores for the 2000-01 MEA in reading, writing, mathematics, science and technology, and social studies were developed by equating the 2000-01 raw scores to the 1999-2000 raw scores. Equating scores from alternate forms of a test adjusts for any difference in difficulty and allows for scores from the different forms to be comparable. Because the 1999-2000 and 2000-01 versions of each test were developed from the same framework, they may be considered alternate forms. Equating test scores makes it possible for the results of the 2000-01 administration to be reported on the same scale that MEA results were reported on the previous year. The process of equating and scaling does not change the rank ordering of students, give more weight to particular questions, or change students' performance level classifications.

Equating for the MEA uses the *anchor-test-nonequivalent-groups* design with external anchor described by Petersen, Kolen, & Hoover (1993). The "anchor test" for reading, mathematics, science and technology, and social studies is a set of matrix items included in both test administrations. These items are external to the test in that they do not contribute to the students' raw scores in either administration of the test. The groups of students who took each test in 1999-2000 and 2000-01 were naturally occurring groups and no assumption was made regarding their equivalence. Item Response Theory (IRT) is particularly useful in this type of equating (Allen & Yen, 1979). All IRT calibrations performed on MEA are for equating.

Equating for MEA writing used the reading scaled scores as the "anchor test." The Tucker Method described in Kolen and Brennan (1995, pp. 105-111) was implemented.

Developing equated scores for the 2000-01 MEA involved several steps. The first step was to construct the "anchor test;" that is, to determine the set of equating items. The second step was to calibrate the items in an IRT model. The IRT model used was a combination of the three-parameter logistic (3PL) model for multiple-choice items, the two-parameter logistic (2PL) model for short-answer items, and the graded response model (GRM) for the constructed-response items. The calibration was first performed on the 1999-2000 data. The item parameters of the equating items resulting from this calibration were fixed for the calibration of the 2001 data. Fixing the parameters

of the equating items ensures that the two "forms" of the test (1999-2000 and 2000-01) were calibrated to the same scale of the trait being measured.

DETERMINING THE SETS OF EQUATING ITEMS

During the development stage of the 2000-01 MEA, matrix items that were also administered in the previous year were identified as potential equating items. These items were designated based on the following criteria:

1. The average difficulty of the equating items was about the same as the average difficulty of the 1999-2000 test.

2. The total points from the equating items were about equivalent to 40% of the total points on the test.

3. The position of each item in the 2000-01 test form was about the same as its position in the 1999-2000 test form.

4. The distribution of the items across different relevant categories (i.e. item types and content areas) was similar to that of the whole test.

5. There should not be any significant change in the item from one administration to the other.

To determine the final set of equating items for each grade level and subject combination, a differential item functioning (DIF) approach using the delta plot method was applied. The p-values of each multiple-choice and short-answer item were transformed to the delta metric. Each item has two p-values—one for each test administration. The delta scale is an inverse normal transformation of percentage correct to a linear scale with a mean of 13 and standard deviation of 4 (Holland & Wainer, 1993). A high delta value indicates a difficult item. For constructed-response items, the average score divided by the maximum possible score or adjusted p-value was transformed to the delta metric. The delta values computed for the potential equating items were plotted for each subject (reading, mathematics, science and technology, social studies, health education and visual and performing arts) in each grade level (4, 8, 11).

Figure 12-1 is an example of delta plot for equating items. The dark diagonal line is the trend line and the light diagonal line is the identity line. Different shapes were used to identify different item types: ♦ for multiple-choice items; ▲ for short-answer items; and, ● for constructed-response items. The perpendicular distance of each item to the regression line was computed. The unshaded shape indicates the item with the greatest perpendicular distance

from the regression line. Items that were not more than three standard deviations away from the regression line were used as equating items.

An additional criterion was also applied for constructed response items in order to be included as equating or anchor items. The average score for each potential equating item should not significantly differ for the two years.

Figure 12-1
Sample Delta Plot
(♦ MC   ▲ SA   ● CR)
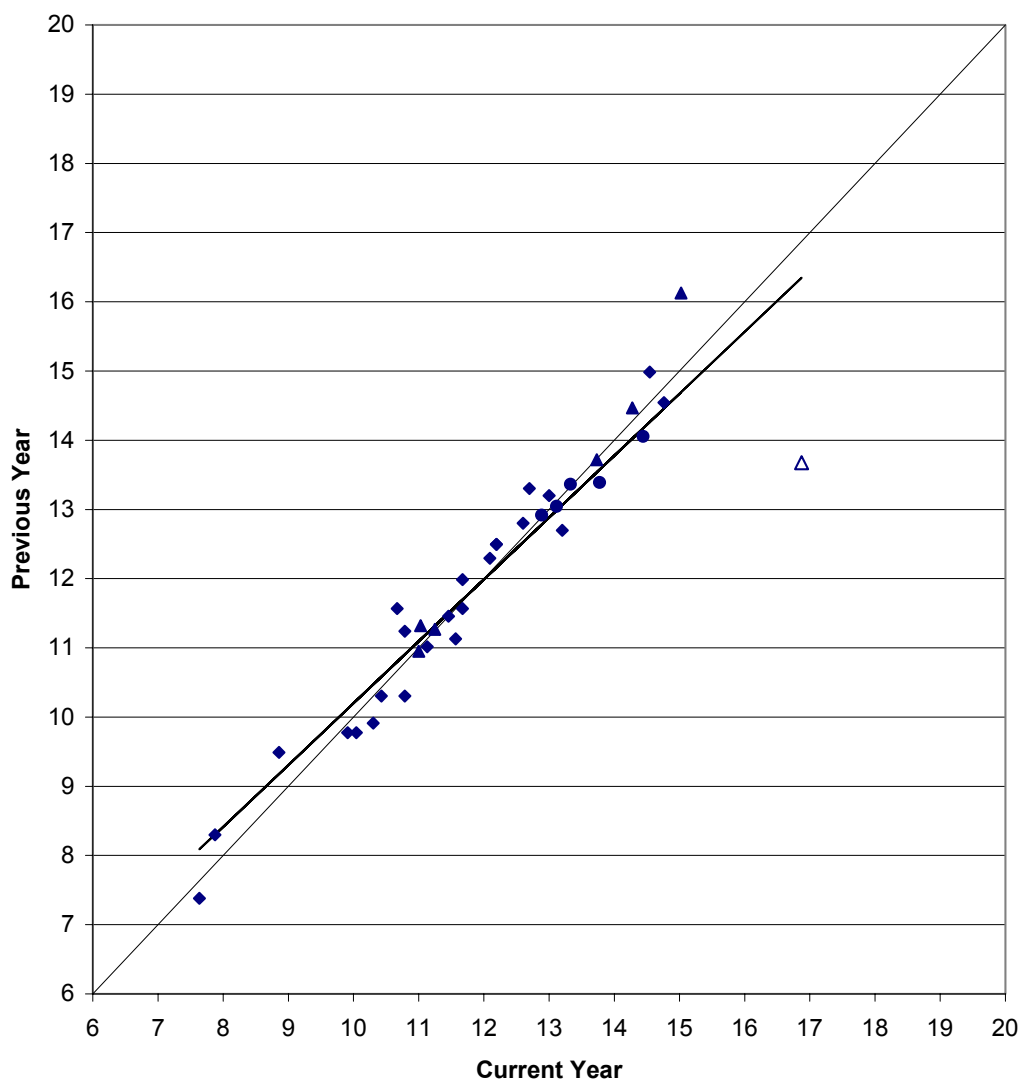
ITEM CALIBRATIONS

IRT calibration was performed on the common and matrix items from the 1999-2000 MEA using a combination of IRT models specific to item types (i.e., 3PL for multiple-choice, 2PL for short-answer, and GRM for constructed response). Each of these models expresses examinees' tendencies to achieve certain scores on the items contributing to a scale as a function of a parameter that is not directly observed and commonly referred to as θ. Using the current version of PARSCALE, item parameters were estimated based on those models.

To calibrate items for 2000-01, parameters for equating items were fixed to their calibrated values from the previous year. This ensured that the tests for the two years were calibrated to the same ability scale. The item parameters resulting from their calibration become the basis for equated scores.

SCORES FOR READING, MATHEMATICS, SCIENCE AND TECHNOLOGY, AND SOCIAL STUDIES

For reading, mathematics, science and technology, and social studies, IRT parameters resulting from the calibrations were used to estimate student abilities. The estimated student abilities are based only on common items. The cumulative distributions of raw scores and scaled scores for each subject and grade combination for 2000-01 and 1999-2000 were used to find the equated cutpoints. Thus, for the 2000-01 MEA a new set of cutpoints was obtained. This process is described using Figure 12-2.

Suppose c_{2000} is a cutpoint resulting from the standard setting in 1999-2000. This cutpoint is in the raw score metric. Using the frequency distribution of the raw scores for 1999-2000, the cumulative percentage associated with this cutpoint was estimated through linear interpolation. The θ value associated with this cumulative percentage was determined using the frequency distribution of ability estimates. Because ability for 1999-2000 and 2000-01 is on the same θ scale, the obtained θ value corresponds to the same ability for both years. The 2000-01 cumulative percentage associated with this θ was then mapped to a 2000-01 raw score through linear interpolation resulting in c_{2001}.

The above process was used for each cutpoint set in 1999-2000 for each grade for reading, mathematics, science and technology, and social studies. The resulting cutpoints are presented in Table 12-1. These cutpoints were

used to obtain new scaling parameters m_1, m_2, b_1, and b_2, which were then used to compute the scaled scores for 2000-01. The new scaling parameters are presented in Table 12-2.

The functions that translate raw scores to scaled scores are:

$$S = m_1 r + b_1 \qquad \text{if } r < P, \text{ and}$$
$$S = m_2 r + b_2 \qquad \text{if } r > P$$

where S is the scaled score, r is the raw score, and P is the threshold for "Meets the Standard."

Table 12-1
Threshold (Minimum) Total Test Score For Each Performance Category for Reading, Mathematics, Science, and Social Studies

Grade	Subject Area	Maximum Score on Test	Threshold Score		
			Exceeds The Standards	Meets The Standards	Partially Meets the Standards
4	Reading	48	42.88	29.46	17.64
	Mathematics	50	43.19	32.04	19.91
	Science and Technology	50	42.82	36.14	21.49
	Social Studies	50	38.29	26.76	14.47
8	Reading	48	42.28	30.20	18.38
	Mathematics	50	46.54	33.00	18.94
	Science and Technology	50	41.35	32.42	20.28
	Social Studies	50	39.98	30.36	18.96
11	Reading	48	43.37	32.12	19.44
	Mathematics	50	42.77	27.45	15.38
	Science and Technology	50	42.19	31.92	15.43
	Social Studies	50	39.84	27.09	16.52

Table 12-2
Transformation Constants Used to Compute Scaled Scores for Reading, Mathematics, Science, and Social Studies

Grade	Subject Area	Transformation Constants			
		m_1	b_1	m_2	b_2
4	Reading	1.69	491.13	1.49	497.13
	Mathematics	1.65	488.16	1.79	483.52
	Science and Technology	1.37	491.65	3.00	432.73
	Social Studies	1.63	497.43	1.73	494.58
8	Reading	1.69	489.90	1.66	490.99
	Mathematics	1.42	494.08	1.48	492.22
	Science and Technology	1.65	487.58	2.24	468.37
	Social Studies	1.75	487.74	2.08	477.82
11	Reading	1.58	490.34	1.78	483.88
	Mathematics	1.66	495.50	1.31	505.17
	Science and Technology	1.21	502.30	1.95	478.81
	Social Studies	1.89	489.76	1.57	498.49

Figure 12-2
Finding Equated Cutpoints
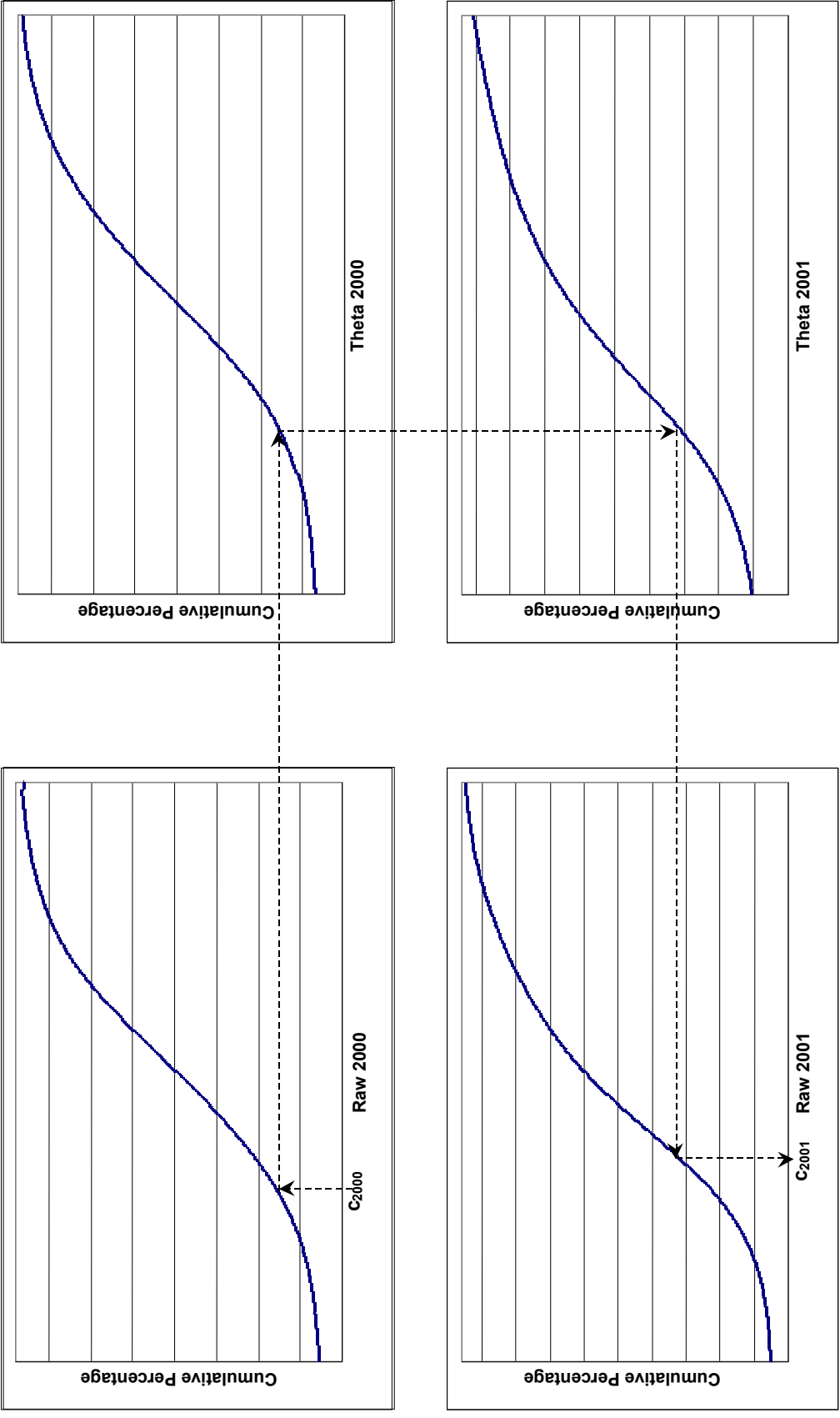
Using the reading scaled scores as the "anchor test," 2000-01 writing raw scores were equated to 1998-99 writing raw scores using the Tucker Method described in Kolen & Brennan (1995, pp. 109-111). The equated scores were then transformed to scaled scores using the linear equations on page 70. The values of ms and bs are on Table 12-3.

SCALED SCORES FOR HEALTH AND VISUAL AND PERFORMING ARTS

The equating procedure for health and visual and performing arts is the same as that for reading, mathematics, science and technology, and social studies. However, the scaled scores for health and visual and performing arts are linear transformations of θ scores and not raw scores like in reading, mathematics, science and technology, and social studies.

The functions that translate θs to scaled scores are

$$S = m_1\theta + b_1 \qquad \text{if } \theta < P, \text{ and}$$
$$S = m_2\theta + b_2 \qquad \text{if } \theta > P$$

where S is the scaled score, θ is the ability estimate, and P is the threshold for "Meets the Standard." These scaling parameters m_1, m_2, b_1, and b_2 are based on the results of standard setting processes implemented for health and visual and performing arts in 1999-2000. These constants are also presented in Table 12-3.

Table 12-3
Transformation Constants Used to Compute Scaled Scores for Writing, Health, and Visual and Performing Arts

Grade	Subject Area	Transformation Constants			
		m_1	b_1	m_2	b_2
4	Writing	2.47	495.08	2.31	498.11
	Health	19.68	533.95	10.13	537.37
	Visual and Performing Arts	8.21	534.14	11.40	531.48
8	Writing	2.19	501.32	2.79	490.60
	Health	12.29	537.45	10.74	537.89
	Visual and Performing Arts	9.39	534.99	14.29	531.86
11	Writing	2.92	482.21	2.49	490.85
	Health	13.89	536.26	10.78	537.32
	Visual and Performing Arts	5.12	536.29	14.81	527.37

CHAPTER 13—ITEM ANALYSES

As noted in Brown (1983), "a test is only as good as the items it contains." A complete evaluation of a test's quality must include an evaluation of each question. Both the *Standards for Educational and Psychological Testing* and the *Code of Fair Testing Practices in Education* include standards for identifying quality questions. Questions should assess only knowledge or skills that are under assessment and should avoid assessing irrelevant factors. They should also be unambiguous and free of grammatical errors, potentially insensitive content or language, and other confounding characteristics. Further, questions must not unfairly disadvantage test takers from particular racial, ethnic, or gender groups.

Both qualitative and quantitative analyses are conducted to ensure that MEA questions meet these standards. Previous sections in this report have delineated the qualitative checks on question quality. The current chapter focuses on more quantitative evaluations. The statistical evaluations are presented in three sections: 1) difficulty indices, 2) item-test correlations, and 3) subgroup differences in item performance. The results presented in this chapter are based on the statewide administrations of the MEA in December of 2000 and March of 2001. About 16,300 grade 4 students, 17,200 grade 8 students, and 15,000 grade 11 students participated in the assessment.

DIFFICULTY INDICES

All multiple-choice, short-answer, and constructed-response questions were evaluated in terms of difficulty and relationship to overall score according to standard classical test theory practice. Difficulty was measured by averaging the proportion of points received across all students who received the question. Multiple-choice and short-answer questions were scored dichotomously (correct v. incorrect), so for these questions the difficulty index is simply the proportion of students who correctly answered the question. Constructed-response questions allowed for scores between zero and four. By computing the difficulty index as the average proportion of points received, the indices for multiple-choice, short-answer, and constructed-response questions are placed on a similar scale; the index ranges from zero to one regardless of the question type. Although this index is traditionally described as a measure of difficulty (as it is described here), it is properly interpreted as an easiness index because larger values indicate

easier questions. An index of zero indicates that no student received credit for the question, and an index of one indicates that every student received full credit for the question.

Questions that are answered correctly by almost all students provide little information about differences in student ability, but they do indicate knowledge or skills that have been mastered by most students. Similarly, questions that are correctly answered by very few students may indicate knowledge or skills that have not yet been mastered by most students, but such questions provide little information about differences in student ability. In general, to provide best measurement, difficulty indices should range from near-chance performance (.25 for four-option, multiple-choice questions or essentially zero for short-answer and constructed-response questions) to .90. Indices outside this range indicate questions that were either too difficult or too easy for the target population.

Although difficulty is an important question characteristic, the relationship between performance on a question and performance on the whole test or a relevant test section may be more critical. A question that assesses relevant knowledge or skills should relate to other questions that are purported to be measuring the same knowledge or skills.

ITEM-TEST CORRELATIONS

Within classical test theory, these relationships are assessed using correlation coefficients that are typically described as either item-test correlations or, more commonly, discrimination indices. The discrimination index used to analyze MEA multiple-choice items and zero- or one-scored short-answer items was the point-biserial correlation between item score and a criterion total score on the test. As such, the index ranges from -1 to 1, with the magnitude and sign of the index indicating the relationship's strength and direction, respectively. For constructed-response items, item discrimination indices were based on the Pearson product-moment correlation. The theoretical range of these statistics is also from -1 to 1, with a typical range from .3 to .6.

In general, discrimination indices are interpreted as indicating the degree to which high- and low-ability students perform differently on a question or, equivalently, the degree to which performance on a question helps to differentiate between high- and low-ability students. From this perspective, indices near 1 indicate that high-ability students are more likely to answer the question correctly, indices near -1 indicate that low-ability students are more

likely to answer the question correctly, and indices near 0 indicate that performance on the question is equally likely to be answered correctly by high- and low-ability students.

Discrimination indices can be thought of as measures of how closely a question assesses the same knowledge and skills assessed by other questions contributing to the criterion total score; that is, the discrimination index can be interpreted as a measure of construct consistency. In light of this interpretation, the selection of an appropriate criterion total score is crucial to the interpretation of the discrimination index. For the 2000-2001 MEA, the criterion score for each common item is the total score for all common items. For each matrix item the criterion score is the total score for the form in which that item is positioned.

SUMMARY OF ITEM ANALYSIS RESULTS

Summary statistics of the difficulty and discrimination indices for each question are provided in Tables 13-1–13-3. In general, the question difficulty and discrimination indices are in acceptable and expected ranges. Very few questions were answered correctly at near-chance or near-perfect rates. Similarly, the positive discrimination indices indicate that most questions were assessing consistent constructs, and students who performed well on individual questions tended to perform well overall. There were a small number of questions with near-zero discrimination indices, but none was reliably negative. Occasionally, questions with less-desirable statistical characteristics need to be included in assessments to ensure that content is appropriately covered, but there were very few such cases.

A comparison of indices across grade levels is complicated because these indices are population dependent. Direct comparisons would require that either the questions or students were common across groups. However, one can say that with respect to multiple-choice items, students in all three grades did similarly in reading, health education, and visual and performing arts. For the remaining three subject areas, students in the eleventh grade had more difficulty answering eleventh-grade questions than students in the fourth and eighth grade had answering questions on the fourth- and eighth-grade tests. For two of these subject areas, mathematics and science and technology, fourth-graders had an easier time answering fourth-grade questions than did eighth-grade students on the eighth-grade test.

Comparing the difficulty indices of multiple-choice and short-answer or constructed-response questions is inappropriate because multiple-choice questions can be answered correctly by guessing. Thus, it is not surprising that the difficulty indices for multiple-choice questions tend to be higher (indicating easier questions) than the difficulty indices for other question types. Similarly, the partial credit allowed by constructed-response questions is advantageous in the computation of question-test correlations, so the discrimination indices for these questions tend to be larger than the discrimination indices of other question types.

Table 13-1
Average Difficulty and Discrimination of Different Item Types For Each Grade-Content Area Combination - Grade 4

Content Area	Statistics	Item Type		
		All	Multiple Choice	Constructed response
Reading	Difficulty	0.55 (0.20)	0.65 (0.15)	0.34 (0.11)
	Discrimination	0.51 (0.11)	0.47 (0.11)	0.57 (0.08)
	N	135	90	45
Mathematics	Difficulty	0.49 (0.21)	0.60 (0.16)	0.31 (0.15)
	Discrimination	0.43 (0.11)	0.39 (0.10)	0.51 (0.08)
	N	126	80	46
Science and Technology	Difficulty	0.53 (0.22)	0.61 (0.18)	0.29 (0.13)
	Discrimination	0.36 (0.10)	0.35 (0.10)	0.40 (0.09)
	N	126	92	34
Social Studies	Difficulty	0.51 (0.19)	0.58 (0.16)	0.31 (0.09)
	Discrimination	0.36 (0.10)	0.34 (0.10)	0.42 (0.07)
	N	126	92	34
Health	Difficulty	0.50 (0.19)	0.57 (0.20)	0.40 (0.11)
	Discrimination	0.33 (0.10)	0.29 (0.09)	0.40 (0.08)
	N	120	72	48
VPA	Difficulty	0.55 (0.17)	0.57 (0.17)	0.43 (0.06)
	Discrimination	0.30 (0.09)	0.29 (0.10)	0.38 (0.03)
	N	84	72	12

Table 13-2
Average Difficulty and Discrimination of Different Item Types For Each Grade-
Content Area Combination - Grade 8

| Content Area | Statistics | Item Type | | Constructed |
		All	Multiple Choice	response
Reading	Difficulty	0.57 (0.19)	0.67 (0.14)	0.37 (0.12)
	Discrimination	0.47 (0.11)	0.42 (0.09)	0.57 (0.08)
	N	135	90	45
Mathematics	Difficulty	0.41 (0.20)	0.51 (0.16)	0.23 (0.10)
	Discrimination	0.46 (0.12)	0.41 (0.10)	0.56 (0.08)
	N	125	80	45
Science and Technology	Difficulty	0.48 (0.21)	0.56 (0.19)	0.27 (0.10)
	Discrimination	0.37 (0.11)	0.34 (0.10)	0.47 (0.10)
	N	126	92	34
Social Studies	Difficulty	0.48 (0.21)	0.56 (0.19)	0.27 (0.10)
	Discrimination	0.40 (0.13)	0.37 (0.12)	0.50 (0.11)
	N	126	92	34
Health	Difficulty	0.55 (0.27)	0.65 (0.18)	0.40 (0.32)
	Discrimination	0.33 (0.15)	0.32 (0.10)	0.34 (0.21)
	N	119	72	47
VPA	Difficulty	0.55 (0.16)	0.57 (0.16)	0.41 (0.05)
	Discrimination	0.34 (0.09)	0.33 (0.09)	0.42 (0.05)
	N	84	72	12

Table 13-3 Average Difficulty and Discrimination of Different Item Types For Each Grade-Content Area Combination – Grade 11				
		Item Type		
Content Area	Statistics	All	Multiple Choice	Constructed response
Reading	Difficulty	0.58 (0.21)	0.68 (0.16)	0.40 (0.15)
	Discrimination	0.52 (0.14)	0.47 (0.14)	0.61 (0.09)
	N	135	90	45
Mathematics	Difficulty	0.34 (0.18)	0.42 (0.15)	0.18 (0.12)
	Discrimination	0.46 (0.13)	0.41 (0.11)	0.56 (0.11)
	N	125	80	45
Science and Technology	Difficulty	0.41 (0.20)	0.48 (0.19)	0.23 (0.11)
	Discrimination	0.38 (0.14)	0.34 (0.12)	0.48 (0.11)
	N	126	92	34
Social Studies	Difficulty	0.42 (0.18)	0.48 (0.16)	0.26 (0.10)
	Discrimination	0.38 (0.15)	0.33 (0.12)	0.53 (0.12)
	N	125	92	33
Health	Difficulty	0.62 (0.20)	0.66 (0.16)	0.53 (0.25)
	Discrimination	0.42 (0.13)	0.37 (0.11)	0.53 (0.11)
	N	111	72	39
VPA	Difficulty	0.54 (0.16)	0.55 (0.17)	0.44 (0.05)
	Discrimination	0.39 (0.11)	0.37 (0.10)	0.51 (0.05)
	N	84	72	12

SUBGROUP DIFFERENCES IN TEST QUESTION PERFORMANCE

The Code of Fair Testing Practices in Education explicitly states that subgroup differences in performance should be examined when sample sizes permit, and actions should be taken to make certain that differences in performance are due to construct-relevant, rather than irrelevant, factors. The *Standards for Educational and Psychological Testing* includes similar guidelines. As part of the effort to identify such problems, MEA questions were evaluated in terms of differential item functioning (DIF) statistics.

DIF procedures are designed to identify questions for which subgroups of interest perform differently beyond the impact of differences in overall achievement. For the MEA, the standardization DIF procedure (Dorans and Kulick, 1986) was employed to evaluate subgroup differences between male and female. This procedure calculates the difference in item performance for groups of students matched for achievement on the total test. That is, the average item performance is calculated for students at every total score; then an overall average is calculated weighting the total score distribution so it is the same for the two groups.

The index ranges from –1 to 1 for multiple-choice and short-answer questions and is adjusted to the same scale for constructed-response questions. Negative numbers indicate that the question was more difficult for female students. Positive numbers indicate that the question was easier for female students.

Dorans and Holland (1993) suggested that index values between –0.05 and 0.05 should be considered negligible for dichotomously scored questions (such as MEA multiple-choice questions). Most MEA questions fall within this range. Dorans and Holland further stated that dichotomously scored questions with values between –0.10 and –0.05 and between 0.05 and 0.10 (i.e., "low" DIF) should be inspected to ensure that no possible effect is overlooked, and that questions with values outside the [–0.10, 0.10] range (i.e., "high" DIF) are more unusual and should be examined very carefully. These standards can be applied to constructed-response questions by accounting for the larger range of possible index values and scaling appropriately. That is, values of the DIF index can range from –4.0 to 4.0, so the corresponding ranges are between –0.2 and 0.2 for negligible difference, between –0.4 and –0.2 and between 0.2 and 0.4 for "low" DIF and outside [-0.4, 0.4] for "high" DIF.

DIF indices indicate differential performance between two groups. That differential performance may or may not be indicative of bias in the test. Course-taking patterns, group differences in interests, or differences in school curricula can lead to DIF. If subgroup differences in performance are related to construct-relevant factors, the questions should be considered for inclusion on a test.

Each question was categorized according to the guidelines adapted from Dorans and Holland (1993). Tables 13-4 to 13-6 provide the number of questions in each of the three DIF categories for male v. female for each grade level tested. There are some MEA questions categorized as "low" or "high" DIF. These indices must not be interpreted as indisputable evidence of bias. Both the *Code of Fair Testing Practices in Education* and the *Standards for Educational and Psychological Testing* assert that test questions must be free from construct-irrelevant sources of differential difficulty. If subgroup differences in performance can be plausibly attributed to construct-relevant factors, the questions may be included on a test. What is important is to determine if the cause of this differential performance is construct relevant.

Table 13-4

Differential Item Functioning (DIF) Categorization Item Type: Grade 4

Content Area	Item Type	Negligible DIF				Low DIF				High DIF			
		Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%
ELA	Multiple Choice	25	48	73	54.10	2	13	15	11.10	0	2	2	1.50
	Constructed Response	24	18	42	31.10	2	1	3	2.20	0	0	0	0.00
Mathematics	Multiple Choice	39	42	81	64.3	3	8	11	8.7	0	0	0	0
	Constructed Response	17	11	28	22.2	5	0	5	4	1	0	1	0.8
Science and Technology	Multiple Choice	19	29	48	57.1	3	19	22	26.2	0	2	2	2.4
	Constructed Response	9	0	9	10.7	2	0	2	2.4	1	0	1	1.2
Social Studies	Multiple Choice	22	34	56	44.8	3	19	22	17.6	0	2	2	1.6
	Constructed Response	27	12	39	31.2	2	4	6	4.8	0	0	0	0
Health	Multiple Choice	16	30	46	38.3	4	17	21	17.5	1	4	5	4.2
	Constructed Response	28	11	39	32.5	5	3	8	6.7	0	1	1	0.8
VPA	Multiple Choice	24	43	67	53.20	7	13	20	15.90	2	3	5	4.00
	Constructed Response	21	7	28	22.20	5	1	6	4.80	0	0	0	0.00

Table 13-5

Differential Item Functioning (DIF) Categorization Item Type: Grade 8

Content Area	Item Type	Negligible DIF				Low DIF				High DIF			
		Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%
ELA	Multiple Choice	22	47	69	51.10	3	15	18	13.30	0	3	3	2.20
	Constructed Response	32	9	41	30.40	1	3	4	3.00	0	0	0	0.00
Mathematics	Multiple Choice	25	31	56	44.8	7	15	22	17.6	0	2	2	1.6
	Constructed Response	26	8	34	27.2	8	1	9	7.2	1	1	2	1.6
Science and Technology	Multiple Choice	25	30	55	43.7	6	21	27	21.4	0	10	10	7.9
	Constructed Response	12	12	24	19	7	3	10	7.9	0	0	0	0
Social Studies	Multiple Choice	28	35	63	50.4	5	21	26	20.8	0	3	3	2.4
	Constructed Response	8	10	18	14.4	11	2	13	10.4	0	2	2	1.6
Health	Multiple Choice	23	30	53	49.1	1	11	12	11.1	0	7	7	6.5
	Constructed Response	21	12	33	30.6	2	1	3	2.8	0	0	0	0
VPA	Multiple Choice	18	22	40	47.60	3	22	25	29.80	1	6	7	8.30
	Constructed Response	8	0	8	9.50	4	0	4	4.80	0	0	0	0.00

Table 13-6

Differential Item Functioning (DIF) Categorization Item Type: Grade 11

Content Area	Item Type	Negligible DIF				Low DIF				High DIF			
		Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%
ELA	Multiple Choice	31	43	74	54.8	2	12	14	10.4	0	2	2	1.5
	Constructed Response	24	9	33	24.4	8	4	12	8.9	0	0	0	0
Mathematics	Multiple Choice	25	31	56	44.8	7	15	22	17.6	0	2	2	1.6
	Constructed Response	26	8	34	27.2	8	1	9	7.2	1	1	2	1.6
Science and Technology	Multiple Choice	25	30	55	43.7	6	21	27	21.4	0	10	10	7.9
	Constructed Response	12	12	24	19	7	3	10	7.9	0	0	0	0
Social Studies	Multiple Choice	28	35	63	50.4	5	21	26	20.8	0	3	3	2.4
	Constructed Response	8	10	18	14.4	11	2	13	10.4	0	2	2	1.6
Health	Multiple Choice	17	27	44	39.6	2	20	22	19.8	1	5	6	5.4
	Constructed Response	23	10	33	29.7	2	2	4	3.6	1	1	2	1.8
VPA	Multiple Choice	18	22	40	47.6	3	22	25	29.8	1	6	7	8.3
	Constructed Response	8	0	8	9.5	4	0	4	4.8	0	0	0	0

CHAPTER 14—RELIABILITY

Although an individual question's performance is an important focus for evaluation, a complete evaluation of an assessment must also address the way that questions function together and complement one another. Any measurement includes some amount of measurement error; that is, no measurement can be perfectly accurate. This is true of academic assessments—no assessment can measure students with perfect accuracy; some students will receive scores that underestimate their true ability, and other students will receive scores that overestimate their true ability. Questions that function well together produce assessments that have less measurement error; that is, the errors made should be small on average. Such assessments are described as reliable.

There are a number of ways to estimate an assessment's reliability. One approach is to split all test questions into two groups and then correlate students' scores on the two half-tests. This is known as a split-half estimate of reliability. If the two half-test scores correlate highly, questions on the two half-tests must be measuring very similar knowledge or skills. This is evidence that the questions complement one another and function well as a group. This also suggests that measurement error will be minimal.

The split-half method requires the psychometrician to select which questions contribute to each half-test score. This decision may have an impact on the resulting correlation. Cronbach (1951) provided a statistic that avoids this concern about the split-half method. Cronbach's α coefficient is an estimate of the average of all possible split-half reliability coefficients.

RELIABILITY AND STANDARD ERRORS OF MEASUREMENT

Table 14-1 presents descriptive statistics, Cronbach's α coefficient, and raw and scaled score standard errors of measurement for each subject separately for each grade level. The reported reliabilities for writing, health education, and visual and performing arts are the averages of the computed Cronbach's α across forms. The low reliability values can be attributed to the lower number of items in each form in those tests.

Note that two scaled-score standard errors of measurement are presented: one for scaled scores below 542 and one for scaled scores of 542 and above. This is because different slopes are used in the linear transformation to scaled scores at these two different parts of the scaled score range.

Table 14-1
Reliabilities, Standard Errors of Measurement and Descriptive Statistics
MEA 2000-2001

Grade	Content Area	n	Min	Max	Mean	S.D.	Rel.	S.E.M.	Scaled Score	
									<542 S.E.M.	>=542 S.E.M.
4	Reading	15,249	1	47	28.58	7.40	0.82	3.11	3.70	2.18
	Writing	15,102	6	30	14.98	4.34	0.61	2.71	4.80	2.73
	Mathematics	15,581	2	48	25.97	8.21	0.83	3.35	4.11	2.43
	Science/Tech	15,704	3	48	25.49	6.72	0.78	3.16	4.01	2.55
	Social Studies	15,693	2	47	22.65	7.09	0.79	3.22	3.53	2.78
	Health*	16,351	0	20	9.84	3.47	0.65	2.04	3.37	4.47
	VPA*	16,356	0	10	5.14	2.07	0.55	1.38	7.20	5.84
8	Reading	16,139	2	46	28.12	7.46	0.82	3.15	3.81	2.26
	Writing	16,068	6	30	17.29	4.39	0.64	2.65	4.39	2.78
	Mathematics	16,142	0	50	23.8	10.55	0.86	3.95	4.03	2.23
	Science/Tech	16,227	1	48	24.98	7.45	0.8	3.32	4.31	2.84
	Social Studies	16,190	1	48	25.26	7.58	0.83	3.13	4.13	2.57
	Health*	17,199	0	20	10.18	3.82	0.61	2.37	3.63	3.09
	VPA*	17,125	0	10	5.05	2.20	0.60	1.39	7.52	5.86
11	Reading	14,212	2	48	32.02	7.08	0.80	3.14	3.66	2.57
	Writing	14,017	6	30	18.10	4.69	0.70	2.58	4.53	3.30
	Mathematics	13,754	0	49	20.06	9.74	0.88	3.37	3.61	2.01
	Science/Tech	13,924	1	46	20.56	7.54	0.79	3.45	3.56	2.17
	Social Studies	13,877	0	49	21.35	7.94	0.79	3.65	5.04	2.63
	Health*	15,318	0	20	10.87	4.21	0.68	2.37	3.68	3.07
	VPA*	14,982	0	10	5.05	2.41	0.65	1.42	7.32	2.97
*The reported reliability is the average reliability across forms.										

STRATIFIED COEFFICIENT α

According to Feldt and Brennan (1989) a prescribed distribution of items over categories (such as different item types) indicates the presumption that at least a small, but important, degree of unique variance is associated with the categories. In contrast, Cronbach's coefficient α is built upon the assumption that there are no such local or clustered dependencies. A stratified version of coefficient α corrects for this problem.

Stratified coefficient α was calculated separately for each common item test and grade level. The stratification was based on item types (multiple-choice v. constructed response). These results are provided in Table 14-2.

Table 14-2 Coefficients α and Stratified α MEA 2000-2001							
Grade	Subject	α	α_{mc}	N_{mc}	α_{cr}	N_{cr}	Stratified α
04	Reading	0.82	0.74	18	0.74	9(30)	0.84
	Mathematics	0.83	0.70	20	0.76	10(30)	0.84
	Social Studies	0.79	0.67	20	0.69	10(30)	0.80
	Science/Tech	0.78	0.70	20	0.63	10(30)	0.79
08	Reading	0.82	0.69	18	0.76	9(30)	0.84
	Mathematics	0.86	0.77	20	0.80	9(30)	0.88
	Social Studies	0.83	0.74	20	0.75	10(30)	0.84
	Science/Tech	0.80	0.69	20	0.70	10(30)	0.81
11	Reading	0.80	0.69	18	0.72	9(30)	0.82
	Mathematics	0.88	0.78	20	0.83	9(30)	0.89
	Social Studies	0.79	0.60	20	0.75	9(30)	0.81
	Science/Tech	0.79	0.57	20	0.75	10(30)	0.81

RELIABILITY OF PERFORMANCE LEVEL CATEGORIZATION

All test scores contain measurement error; thus classifications based on test scores are also subject to measurement error. After the performance levels were specified and students were classified into those levels, empirical analyses were conducted to determine the statistical accuracy and consistency of the classifications.

ACCURACY

Accuracy refers to the extent to which decisions based on test scores match decisions that would have been made if the scores did not contain any measurement error. Accuracy must be estimated because errorless test scores do not exist.

CONSISTENCY

Consistency measures the extent to which classification decisions based on test scores match the decisions based on scores from a second parallel form of the same test. Consistency can be evaluated directly from actual responses to test questions if two complete and parallel forms of the test are given to the same group of students. This is usually impractical, especially on lengthy tests such as the MEA. To overcome this issue, techniques have been developed to estimate both accuracy and consistency of classification decisions based on a single administration of a test. The technique developed by Livingston and Lewis (1995) was used for the MEA because their technique can be used with both constructed-response and multiple-choice questions.

CALCULATING ACCURACY

All of the accuracy and consistency estimation techniques described below make use of the concept of "true scores" in the sense of classical test theory. A true score is the score that would be obtained on a test that had no measurement error. It is a theoretical concept that cannot be observed, although it can be estimated. Following Livingston and Lewis (1995), the true score distribution for the MEA was estimated using a four-parameter beta distribution, which is a flexible model that allows for extreme degrees of skewness in test scores.

In the Livingston and Lewis method, the estimated "true scores" are used to classify students into their "true" performance category, which is labeled "true status." After various technical adjustments (which are described in Livingston and Lewis, 1995), a 4 × 4 contingency table is created for each test and grade level. The cells in the table are the proportion of students who were classified into each performance category by the actual (or observed) scores on the MEA (i.e., observed status) and by the "true scores" (i.e., "true status"). As an example, Table 14-3 shows the accuracy contingency table for fourth-grade science and technology. The accuracy contingency tables for all grades and subjects are provided in Appendix A (Step 5). Additional steps in the analysis are also shown in Appendix A.

Table 14-3 Accuracy Contingency Table for Grade 4 Science and Technology				
True Status	Observed Status			
	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards
Does Not Meet the Standards	**0.21**	0.06	0.00	0.00
Partially Meets the Standards	0.07	**0.60**	0.02	0.00
Meets the Standards	0.00	0.02	**0.02**	0.00
Exceeds the Standards	0.00	0.00	0.00	**0.00**

Proportions on the diagonal (in bold) indicate exact agreement between the observed status and "true status." If the test were perfectly accurate, all of the off-diagonal cells would be zero. Accuracy is the sum of the diagonal (i.e., the proportion of exact agreement across the four performance levels). In Table 14-3, the diagonal sums to .83, indicating that 83 percent of the students were classified into exactly the same performance categories by their observed scores and their "true scores."

CALCULATING CONSISTENCY

To estimate consistency, the "true scores" are used to estimate the distribution of classifications on an independent parallel test form. After statistical adjustments (see Livingston and Lewis, 1995), a new 4 × 4 contingency table was created for each test and grade level that showed the proportions of students who were classified into each performance category by the actual test and by another (hypothetical) parallel test form. Consistency, which is the proportion of students classified into exactly the same categories by the two forms of the test, is the sum of the diagonal for the new contingency table. The consistency contingency tables are shown under step 7 in Appendix A.

KAPPA

Another way to measure consistency is to use Cohen's (1960) coefficient κ (kappa), which assesses the proportion of consistent classifications after removing the proportion of consistent classification that would be expected by chance. Cohen's κ can be used to estimate the classification consistency of a test from two parallel forms of the test. The second form in this case was the one estimated using the Livingston and Lewis (1995) method. Cohen's κ is shown in Table 14-4. Because κ is corrected for chance, the values of κ are lower than the other consistency estimates in Table 14-4.

RESULTS OF ACCURACY, CONSISTENCY, AND KAPPA ANALYSES

The accuracy, consistency, and kappa indices for all grades and subjects are summarized in Table 14-4.

Table 14-4
Estimates of Accuracy and Consistency of Performance Level Classification

Grade	Subject	Accuracy	Consistency	Kappa (κ)
4	Reading	0.77	0.69	0.47
	Writing	0.75	0.66	0.28
	Mathematics	0.78	0.69	0.50
	Science and Technology	0.83	0.76	0.49
	Social Studies	0.78	0.69	0.46
	Health	0.79	0.69	0.36
	Visual and Performing Arts	0.61	0.49	0.22
8	Reading	0.78	0.69	0.48
	Writing	0.75	0.65	0.38
	Mathematics	0.79	0.71	0.54
	Science and Technology	0.77	0.67	0.46
	Social Studies	0.78	0.69	0.49
	Health	0.77	0.68	0.38
	Visual and Performing Arts	0.58	0.46	0.22
11	Reading	0.77	0.69	0.46
	Writing	0.72	0.62	0.36
	Mathematics	0.81	0.74	0.60
	Science and Technology	0.82	0.73	0.48
	Social Studies	0.75	0.64	0.44
	Health	0.79	0.70	0.39
	Visual and Performing Arts	0.64	0.54	0.30

For certain decisions, concern may be highest for those made about a particular threshold. For example, if a college gave credit to students who achieved an Advanced Placement test score of four or five, but not one, two, or three, one might be interested in the accuracy of the dichotomous decision, below four versus four or above. Table 14-5 reports accuracy and consistency for various dichotomous categorizations on the MEA. MEA partially meets/meets the standards cut accuracy ranges from .79 to .96, and meets/exceeds the standards cut accuracy ranges from .80 to .97. These are relatively high values compared to the 1999 Advanced Placement (AP) accuracy of decisions based on the 2-3 cut and 3-4 cut that ranges from .84 to .95.

Table 14-5
Accuracy and Consistency of Dichotomous Categorizations

Grade	Subject	Accuracy			Consistency		
		D/P*	P/M*	M/E*	D/P	P/M	M/E
4	Reading	0.92	0.87	0.98	0.90	0.81	0.97
	Writing	0.84	0.91	0.99+	0.78	0.87	0.99+
	Mathematics	0.90	0.89	0.99	0.86	0.85	0.99
	Science and Technology	0.87	0.96	0.99+	0.82	0.94	0.99+
	Social Studies	0.92	0.87	0.99	0.88	0.82	0.99
	Health	0.97	0.83	0.98	0.96	0.75	0.97
	Visual and Performing Arts	0.83	0.80	0.96	0.75	0.74	0.94
8	Reading	0.91	0.88	0.98	0.89	0.83	0.98
	Writing	0.90	0.86	0.99+	0.85	0.80	0.99
	Mathematics	0.90	0.90	0.99	0.85	0.87	0.98
	Science and Technology	0.87	0.91	0.99	0.81	0.87	0.99
	Social Studies	0.91	0.89	0.99	0.86	0.85	0.98
	Health	0.97	0.80	0.99	0.96	0.73	0.99
	Visual and Performing Arts	0.81	0.79	0.96	0.73	0.71	0.93
11	Reading	0.95	0.85	0.98	0.94	0.79	0.97
	Writing	0.92	0.81	0.99	0.89	0.75	0.98
	Mathematics	0.91	0.91	0.99	0.87	0.88	0.99
	Science and Technology	0.87	0.95	0.99+	0.80	0.93	0.99+
	Social Studies	0.88	0.88	0.99	0.82	0.84	0.98
	Health	0.96	0.83	0.99	0.94	0.77	0.99
	Visual and Performing Arts	0.80	0.84	0.99	0.74	0.80	0.96

*D/P = Does not meet/partially meets the standards
*P/M = Partially meets/meets the standards
*M/E = Meets/exceeds the standards

CHAPTER 15—VALIDITY

As noted in the *Standards for Educational and Psychological Testing*, validity is the most important consideration in test evaluation. Validity refers to whether specific inferences made from test scores are appropriate, meaningful, and useful. There are several types of validity-related evidence that can be used to support appropriate, meaningful, and useful inferences based on test scores.

CONTENT-RELATED EVIDENCE

As noted in the *Standards*, evidence of test validity begins with test development and continues throughout the entire testing process. Chapters 2 through 9 provide evidence regarding the alignment between the content of the MEA and Maine's *Learning Results*.

EXTERNAL EVIDENCE

External validity of the MEA is conveyed by the relationship of test scores and situational variables such as school transience, course-taking pattern, attitude towards subject matter, and self-image. These situational variables were all based on student questionnaire data collected during the administration of the MEA. Note that not all of the questionnaire items referred to in the following subsections were asked regarding all of the subjects assessed by the MEA. Note also that no inferential statistics are included. However, because the numbers of students are large enough, differences in average scores could be shown to be statistically significant.

SCHOOL TRANSIENCE

This is an evaluation of how time spent in a single school is related to test scores. Students were asked, "In what grade did you start coming to school in this school district?" Medsker (1998) found that typically, students who change schools often do not perform as well as students who regularly attend a single school or school system. Charts in Figure 15-1 clearly indicate that students who spent more time in a single school tended to have higher test scores in reading, science and technology, and visual and performing arts.

Figure 15-1
School Transience and MEA Scores

Question: In what grade did you start coming to school in this school district?

Grade

MEA Average Scores

	Pre-K or Kindergarten	First or Second	Third	Fourth
Mathematics				
Science				
Social Studies				
VPA				

Grade

MEA Average Scores

	Pre-K or Kindergarten	Grade 1-4	Grade 5-7	Grade 8
Mathematics				
Science				
Social Studies				
VPA				
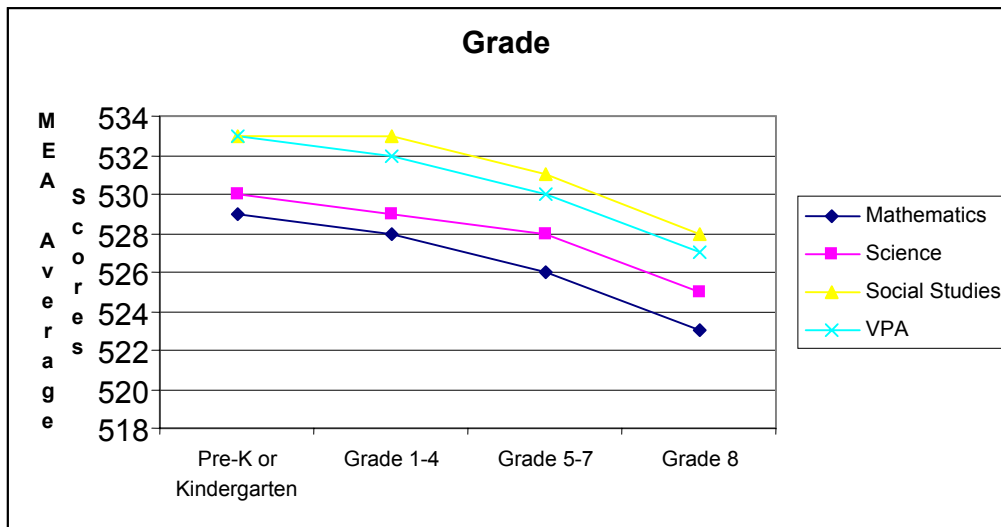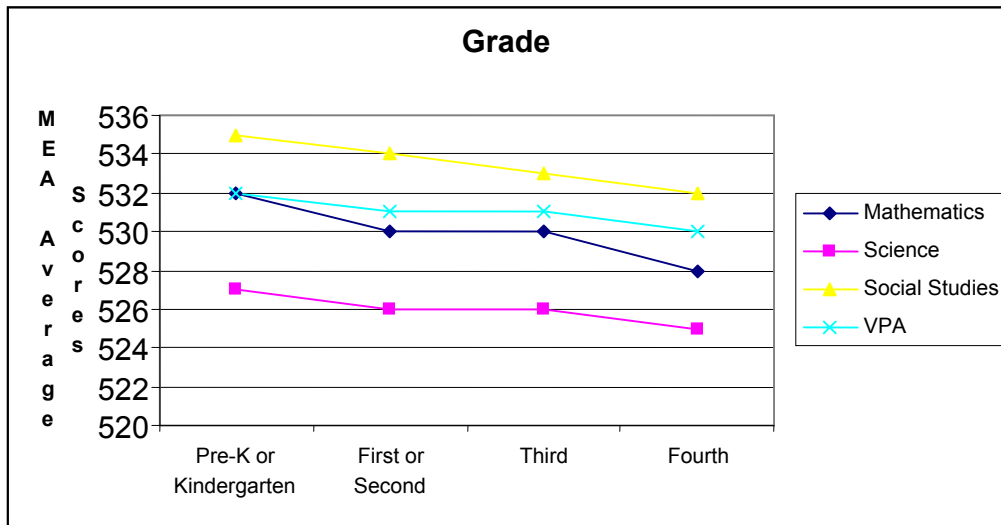
COURSE-TAKING PATTERN

Grade 8 and 11 examinees were asked questions related to their course-taking patterns in mathematics. Eighth graders were asked, "What best describes the mathematics class you are taking in the eighth grade?" and eleventh graders were asked, "What mathematics courses will you complete before you graduate?" Charts in Figure 15-2 both show that higher level mathematics courses are associated with higher MEA mathematics scores.

Figure 15-2
MEA Mathematics Scores and Course-Taking Patterns

Grade 8 Question: What best describes the mathematics class you are taking in the eighth grade?
Grade 11 Question: What mathematics courses will you complete before you graduate?
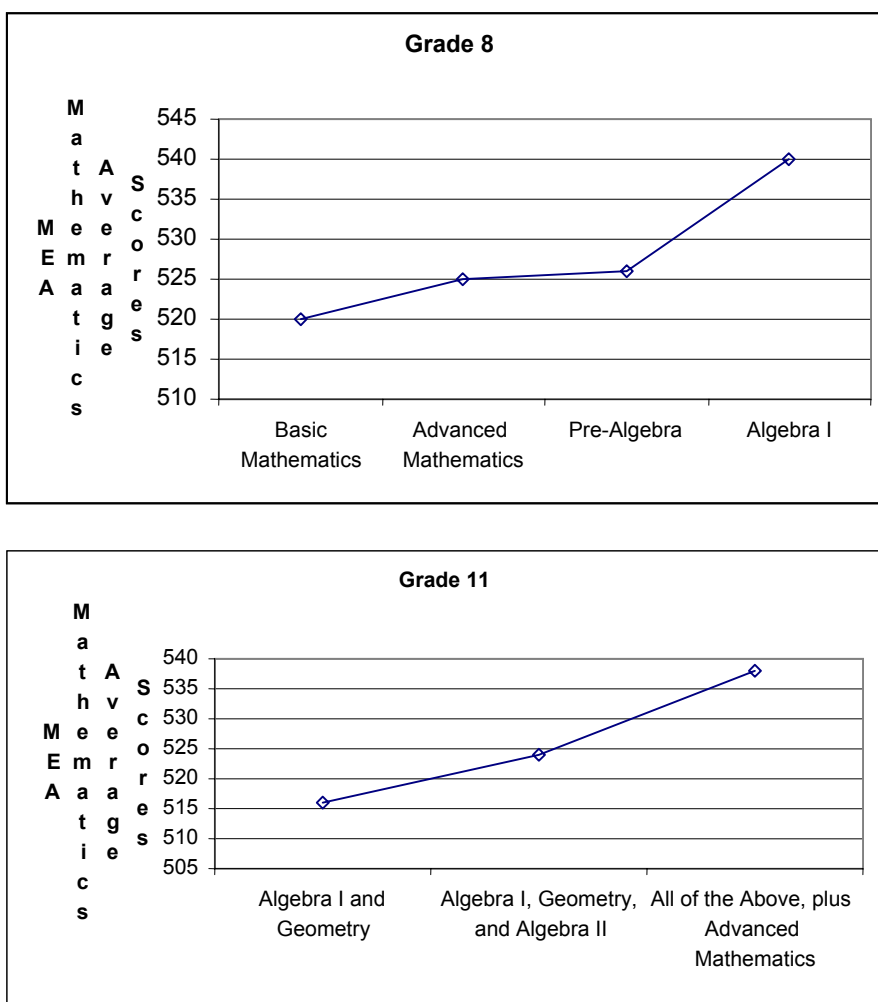
Questionnaire items related to examinees' attitudes toward different subjects tested in the MEA were administered to eighth and eleventh graders. For reading, writing, mathematics, science and technology, social studies, and visual and performing arts, students were asked how they feel about the statement, "My knowledge of [subject] will be useful to me in my future work." For health, students were asked how they feel about the statement, "My knowledge about health education will be helpful to me as an adult." Charts in Figures 15-3 show that students' degree of agreement with statements that indicate their attitudes toward the subjects tested in the MEA are related positively with MEA scores.

Figure 15-3
Attitude Towards Subject Matters and MEA Scores

Question: My knowledge of [subject] will be useful to me [in my future work/as an adult].
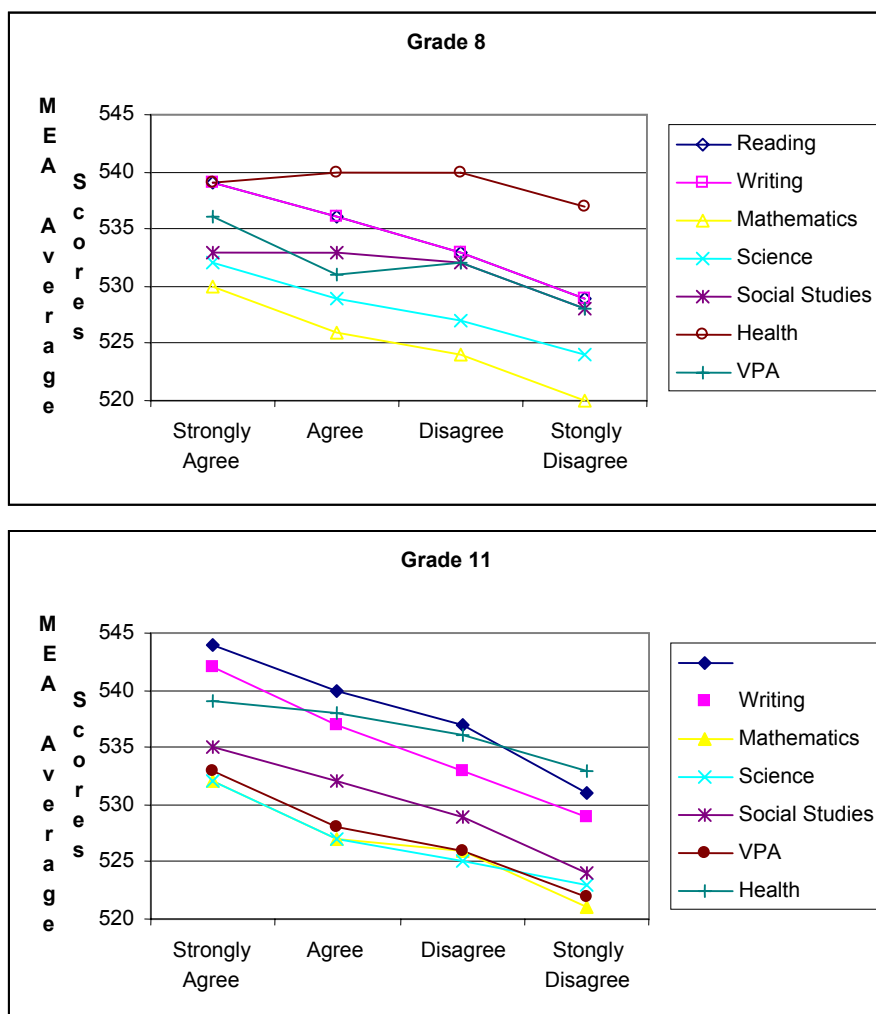
All students who participated in the MEA were asked, "How good are you at reading?" Grade 8 and 11 students were also asked, "How good are you at writing?" Figure 15-4 indicates that there is a positive relationship between students' self-image and their MEA scores in reading and writing.

Figure 15-4
Self-Image and MEA Scores

Question: How good are you at reading/writing?
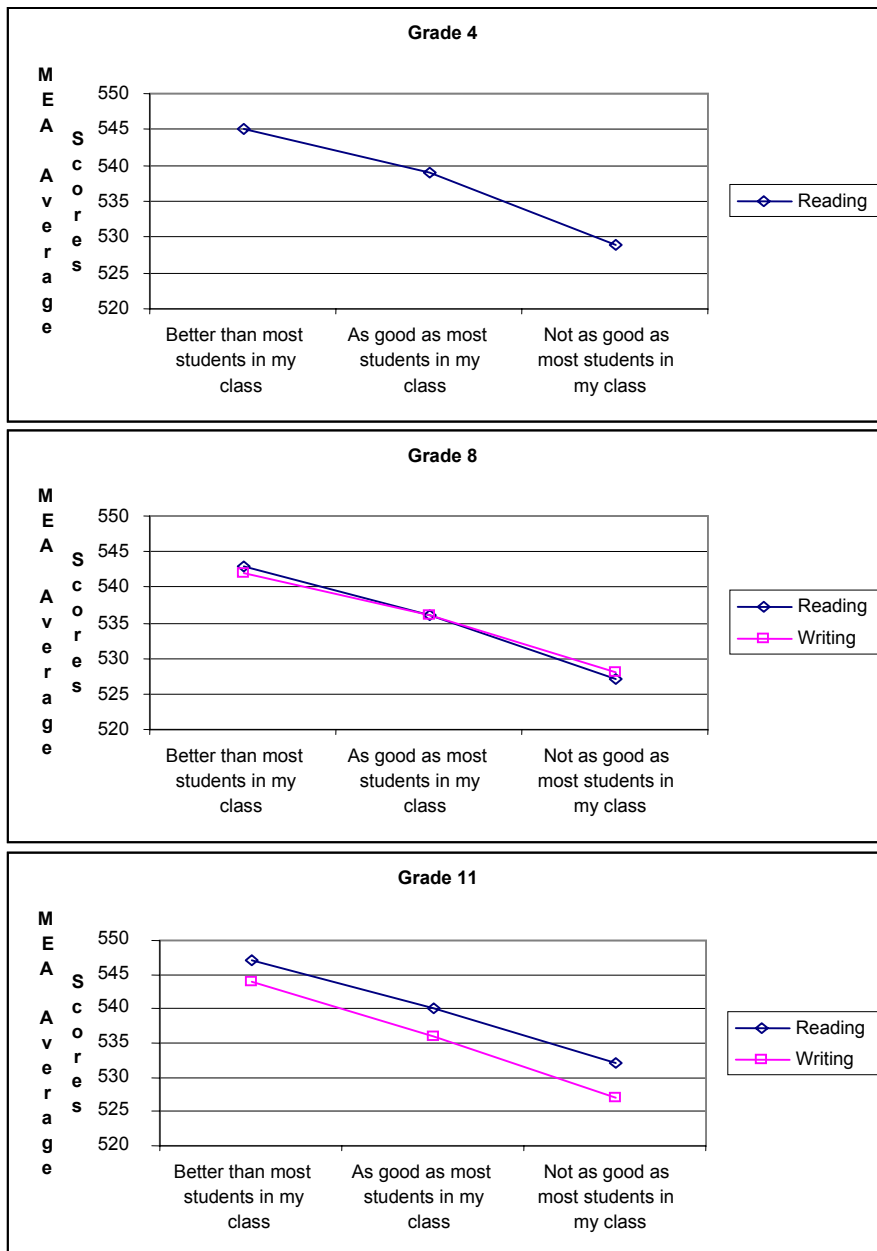
CHAPTER 16—SCORE REPORTING

Table 16-1 lists the primary MEA reports.

<table>
<tr><td colspan="2" align="center">Table 16-1
Primary MEA Reports</td></tr>
<tr><td>1.</td><td>Student Report for Parents/Guardians</td></tr>
<tr><td>2.</td><td>Student Labels</td></tr>
<tr><td>3.</td><td>School Common Item Level Class Report</td></tr>
<tr><td>4.</td><td>School Report</td></tr>
<tr><td>5.</td><td>District Report</td></tr>
<tr><td>6.</td><td>Student Writing CD</td></tr>
</table>

STUDENT REPORT FOR PARENTS/GUARDIANS

Student reports show the scaled score for each subject area, as well as a score band that indicates the standard error of measurement surrounding each score. Performance level definitions are provided so that parents/guardians will understand how to interpret the scaled scores. Specific comments are provided about the student's writing performance. Information is also provided to show how the student's performance compared to the average scores from the student's school, district, and state. An overview of test content is provided, along with a cautionary statement about interpreting scores and guidelines for parents/guardians for helping their children improve.

STUDENT LABELS

To aid schools in keeping track of student scores, schools were supplied with student score information on individual labels that they could affix to files, if desired.

SCHOOL COMMON ITEM LEVEL CLASS REPORT

The *Common Item Level Class Report* shows the answers that each student gave on the common (administered to all students) multiple-choice questions, as well as his/her score on each common constructed-response question. The report also summarizes overall performance at the school, district, and state levels for each of the question types.

SCHOOL AND DISTRICT REPORTS

The school and district reports are intended for administrators and other interested parties. The school report includes performance level definitions, scaled score intervals, and information about how summary statistics are affected by students not tested, all of which are intended to help the reader interpret the report. The school report provides all results for the school, the district, and the entire state. The results provided are

- the number of students tested by student status (regular, students with disabilities, and limited English proficient students) for all subject areas combined and separately for each subject area,

- the percentage of students in each performance level by subject area,

- the distribution of scaled scores by subject area,

- the number of students in each performance level by subject area and student status,

- subject area subscores outlining the number of possible points by learning results standards,

- three-year comparisons of school results, and

- average subject score by number of years in the school or district.

The district report is the same as the school report, except that it does not include the school-level data and the three-year comparisons are by district rather than by school.

STUDENT WRITING CD

The student writing CD contains all student writing responses for each school. The schools are then able to printout and/or review the actual student's work.

Sample reports can be found in Appendix B.

QUALITY CONTROL (QC) PROCESS FOR ENSURING ACCURACY OF PRINTED REPORTS

GENERAL

1. Whenever new reports are received from Measurement, Design, and Analysis (MDA), the *date and time* they were received is written at the top of each report, so that it will be easy to identify the most recent version of each report.

2. For each of the items that follows, a checkmark was put in a logical position on each report to indicate that each check was done. For instance, after verifying that a name is correct, a checkmark is placed next to the name; after verifying that a score is correct or that a bar is the correct length, a checkmark is placed next to it; and so on. This lets other QC staff verify which checks have been done and which have not.

3. When all checks are completed on a given report, the QC staff's initials and that day's date are written at the top of the page so that everyone knows who checked them.

PARENT REPORTS:

Letter Side:

1. Proofing text and formatting of entire side is done once thoroughly, and then spot-checked in additional QC runs.

2. The State MEA Summary Results (bottom right box): the percentages are verified for match with those on the school and district reports for the state (page 2 bar graphs and the results page for each content area). The bars are then checked to make sure that they accurately represent the percentages reported.

Performance Assessment Side:

1. Proofing text and formatting of the entire side is done once thoroughly, and then spot-checked in additional QC runs.

2. It is verified that the student name and grade are the same as those printed on the letter side.

3. Quality control staff also checks to make sure that the performance level corresponds to the scaled score.

4. They verify that the diamond placement in the top box corresponds to score and performance levels and that the range bar does not fall outside of the scale area.

5. If the student was excluded or testing was incomplete, it was verified that no scaled score or performance level appeared, nor were there diamonds or range bars. Instead, it was verified that the words "*excluded*" or "*testing incomplete*" appeared in the performance level box.

6. The performance level and scaled score was compared to the common item report to ensure that they matched. They were also compared to the labels to ensure that they matched.

PARENT REPORTS

Performance Assessment Side:

1. It was verified that the school, district, and state averages matched those in the school and district reports (page 2). The accuracy of the height of the bars was also verified. To make sure that the height of each bar reflected the number on top of the bar, QC staff looked to the left of the bars at the scale. (The bar height should match the performance level.)

2. It was verified that there were no student bars if a student was excluded or testing was incomplete. (Instead, he/she would get the school, district, and state bars only.)

3. Writing comments were checked to verify that the commendations/needs corresponded to the comment codes on the Common Item Class Report (for individual students). It was also checked that the comments were properly categorized (e.g., needs statements into Needs box and commendations statements into Commendation box).

4. Students marked as **NT** (not tested) or **TI** (tested incomplete) *may* still have comments. It was verified that any comments matched what was reported on the Common Item Class Report.

5. Student's Performance in Content Area subcategories: Diamond placement was verified. It was checked that the Diamonds did *not* overlap borders, nor did their corners get cut off. It was also checked that there was a diamond for each of the three categories. If a student was *excluded* or the testing was *incomplete*, then there were *no* diamonds. (If a student had scores for writing but was incomplete or excluded for reading, then the diagram would show two diamonds in the writing category, but no diamond for reading).

LABELS:

1. Spelling, punctuation, and formatting (for margins, fit of text on the label, and so on) were checked.

2. It was verified that the school and district information was correct.

3. The names, proficiency levels, and scaled scores were checked to make sure they matched what was reported on the common item and parent reports.

4. It was also verified that the students listed as belonging to a given school were the same on all reports for that school.

5. The grade was verified and each page of labels was checked to be sure it included information for only *one* school.

COMMON ITEM CLASS REPORTS (READING AND WRITING):

1. The QC staff was directed to proof the text and formatting of the report, including legends (on reverse side), if provided.

2. They also compared the heading information to the shells to be sure that the data in the heading matched the data in the shells.

3. They then verified that the names appear in alphabetical order, and in groups of five.

4. The staff was then told to **highlight** the information for any student who was *excluded or incomplete* (marked with **asterisks**). They subtracted these students from the total and indicated the new total next to the original "group size" in the box at the top of the page. (This is the number used when calculating averages.)

5. It was verified that the number of points per score did not exceed the maximum value indicated in the heading. (If the number 4 is written in the total possible points box, then no one should have an 8 for a score.)

6. The keys were then verified by comparing each correct answer to the incorrect answers listed underneath each question. (For example, if A is the correct key, there should be no As for incorrect answers.)

7. Next, the number of students receiving each *type* of annotation was counted. A need or commendation with the same first letter should only be counted once per student. (For example, a student who received two *needs* that began with a T [for "Topic Development"] and one *commendation* that began with a T, would only be counted

once for the *needs* and *once* for the *commendation*.) These numbers should match those reported on page 8 of the school and district reports.

8. Then the QC staff calculated the average scaled score and the average points earned for the school.

9. Finally, the match to school and district reports took place by adding across classes to get school scores, and across schools to get district scores (remembering to skip the highlighted students and divide by the adjusted group size).

$$\frac{\text{Total of all scaled scores}}{\text{Total number of students}} = \text{average scaled score for the class}$$

$$\frac{\text{Total of all points earned}}{\text{Total number of students}} = \text{average points earned for the class}$$

SCHOOL AND DISTRICT REPORTS:

Page 1:

1. The entire page was proofed for both text and formatting errors, including verifying the page references in the table of contents.

Page 2:

1. The entire page was proofed for both text and formatting errors once thoroughly.

2. It was verified that the scaled scores matched the ones on the parent report and the state-score handout (provided by MDA).

3. The percentage tables were then checked to make sure that the state percentages matched those on the parent reports and handout. The school and district should match the percentages on page 4, 6, or 9.

4. The scores reported for the school and district under Average Performance Score were compared to the averages calculated from the common item reports.

5. Then the staff calculated and verified the accuracy of the Cum. Avg. under Average Performance Score by totaling both averages for the previous year and this year and dividing by two.

6. Finally, they compared this year's report to last year's reports to verify historic data.

Page 3:

1. The entire page was proofed for both text and formatting errors. The informational paragraph at the top of the page was checked so that it refers to school or district as appropriate.

2. It was verified that the students enrolled on the school report and district report equaled the number(s) listed as group size on the common item reports.

3. All percentages (except the last two rows) were computed by taking the number in each row and dividing it by the number enrolled.

Pages 4, 6, and 9 (Reading, Writing, and Health Education) and pages 4, 6, 8,and 11(Mathematics, Science & Technology, Social Studies, and Visual and Performing Arts):

1. All pages were proofread for both text and formatting errors once thoroughly.

2. Quality control staff added up the number of students at each performance level (school "N" and district "N") to get the total included for that content area. It was verified that this total matched the number of students on the common item report (the modified total, minus excluded and testing incomplete students).

3. Then the percent of students at each level was verified by dividing the number at that level by the total number of students included for that content area. The percents were added down the levels to make sure they equaled 99-101.

4. Under "Average Points Attained," the percentage for school, district, and state was verified by dividing the number ("N") by the number of points possible. (Note: If the school or district is small, *some of these cells may be blank*. Each Learning Results Content Standard must have *at least 5 student responses* to be reported in this table.)

Pages 5, 7, and 10 (Reading, Writing, and Health Education) and pages 5, 7, 9,and 11(Mathematics, Science & Technology, Social Studies, and Visual and Performing Arts):

1. All pages were proofread for both text and formatting errors once thoroughly.

2. It was verified that the percentages for each option equaled 99-101 per question.

3. Quality control staff then checked percentages for reasonableness. (If the total *number of students* in a category is *less than 5*, no percentage will be reported. Percentages for "special" categories, such as "Migrant," might total less than 100.)

Page 8: Summary of Annotations Table (Writing Only):

1. The entire page was proofread for both text and formatting errors once thoroughly.

2. It was verified that the number of students receiving a commendation or need matched the number counted on the common item report. A need or commendation with the same first letter was only counted once per student. [For example, a student who received two *needs* that began with a T (for "Topic Development") and one *commendation* that began with a T, would only be counted *once* for the *needs* and *once* for the *commendation*.]

3. Staff then recalculated the percentages by dividing the number of students reported in this table by the total number of students tested in writing for the school and/or district. To get the number tested in writing, the checker added up the number of students at each performance level on page six.

REPORTING IRREGULARITIES

There were two reporting irregularities, both of which occurred in the March 2001 reporting cycle. In the first instance, a problem was detected in the item category information on the Common Item Class Report for all grades and contents, although it is important to note that all student data were correct. The nature of the problem was that some of the performance indicator codes were incorrect (for example, the "2" in a code such as J2 in Science and Technology). Measured Progress corrected the affected codes and shipped revised reports to all schools, along with a cover letter that explained the problem and showed the correction with an explanatory graphic.

In the second instance, a printing problem was detected in the grade 11 Title 1 mathematics report. The data in the "#Tested" and "Average Scaled Score" columns of the school section were reversed in two rows. All performance data on the report were correct, however. Again, Measured Progress produced revised Title 1 reports and shipped them to all grade 11 schools.

REFERENCES

- Allen, Mary J. & Yen, Wendy M. (1979). *Introduction to Measurement Theory.* Belmont, CA: Wadsworth, Inc.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological testing.* Washington, DC: APA.

- Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd Edition). Fort Worth: Holt, Rinehart and Winston.

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description. In P. W. Holland & H. Wainer (Eds.) *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23, 355-368.

- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.) *Educational Measurement* (3rd Ed. Pp. 105-146).

- Holland, P.W., & Wainer, H. (1993). *Differential item functioning.* Hillsdale, NJ: Erlbaum.

- Kolen, M.J., & Brennan, R.L. (1995). *Test equating methods and practices.* New York: Springer-Verlag.

- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179-197.

- Lord, F.M. (1980). A*pplications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

- Medsker, G. J. (1998*). Determining the relationship between student transience and KIRIS school results: are schools with transient students unfairly impacted?* (HumRRO Report, FR-WATSD-98-12). Alexandria, VA: Human Resources Research Organization.

- Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989) Scaling, norming, and equating. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221-262).

APPENDIX A

ACCURACY AND CONSISTENCY OF CLASSIFICATIONS

Accuracy and Consistency of Classifications
Grade 4 Reading

```
                              Step 4


                      Predicted Classification X(1)

                 Does Not   Partially
                 Meet the   Meets the   Meets the      Exceeds
                 the
tstat            Standards  Standards   Standards   Standards   ||   Marginal

Does Not Meet the Standards   0.000928   0.00057   0.00000   0.000000  ||   0.00150
Partially Meets the Standards 0.081268   0.52148   0.05703   0.000001  ||   0.65979
Meets the Standards           0.000012   0.05843   0.23117   0.009890  ||   0.29950
Exceeds the Standards         0.000000   0.00000   0.00990   0.029228  ||   0.03913
                              =========  ========= =========  =========      ========
                              0.082208   0.58048   0.29810   0.039119       0.99992


                              Step 5


                      Actual Classification X(0)

                 Does Not   Partially
                 Meet the   Meets the   Meets the      Exceeds
                 the
tstat            Standards  Standards   Standards   Standards   ||   Marginal

Does Not Meet the Standards   0.00092    0.00042   7.72E-8    806E-21   ||   0.00134
Partially Meets the Standards 0.08062    0.38763   0.09108    2.75E-7   ||   0.55933
Meets the Standards           0.00001    0.04343   0.3692     0.00275   ||   0.41539
Exceeds the Standards         164E-18    2.43E-7   0.01582    0.00812   ||   0.02394
                              =======    =======   =======    =======   ||   =======
Marginal                      0.08155    0.43148   0.4761     0.01087   ||         1


                 Accuracy      Cut #1     Cut #2     Cut #3
                 0.76587      0.91895    0.86547    0.98143


                              Step 6


                                 X(1)

                 Does Not   Partially
                 Meet the   Meets the   Meets the      Exceeds
                 the
tstat            Standards  Standards   Standards   Standards   ||   Marginal

Does Not Meet the Standards   0.026333   0.05540   0.00048   0.000000  ||   0.0822
Partially Meets the Standards 0.055405   0.44513   0.07986   0.000070  ||   0.5805
Meets the Standards           0.000478   0.07986   0.20413   0.013626  ||   0.2981
Exceeds the Standards         0.000000   0.00007   0.01363   0.025425  ||   0.0391
                              =========  ========= =========  =========      ========
                              0.082216   0.58047   0.29810   0.039121       1.0000


                              Step 7


                                 X(0)

                 Does Not   Partially
                 Meet the   Meets the   Meets the      Exceeds
                 the
tstat            Standards  Standards   Standards   Standards   ||   Marginal

Does Not Meet the Standards   0.026119   0.04118   0.00076   0.000000  ||   0.06807
Partially Meets the Standards 0.054947   0.33087   0.12753   0.000020  ||   0.51340
Meets the Standards           0.000474   0.05936   0.32599   0.003787  ||   0.38965
Exceeds the Standards         0.000000   0.00005   0.02176   0.007066  ||   0.02888
                              =========  ========= =========  =========      ========
                              0.081540   0.43146   0.47604   0.010872       1.00000


          Consistency    Cut #1     Cut #2     Cut #3    ||    kappa
            0.69010      0.90263    0.81178    0.97438   ||   0.47214
```

Accuracy and Consistency of Classifications
Grade 4 Writing

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.10135	0.05279	0.000018	3.9191E-13	||	0.15414
Partially Meets the Standards	0.10080	0.62427	0.064865	.000012329	||	0.79004
Meets the Standards	0.00001	0.02084	0.034790	.000142545	||	0.05579
Exceeds the Standards	0.00000	0.00000	0.000000	0	||	0.00000
	=========	=========	=========	==========	||	========
	0.20216	0.69789	0.099673	.000154873		0.99997

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.10834	0.05094	0.00002	678E-15	||	0.1593
Partially Meets the Standards	0.10776	0.60243	0.07167	0.00002	||	0.78189
Meets the Standards	0.00002	0.0201	0.03844	0.00025	||	0.05881
Exceeds the Standards	0	0	0	0	||	0
	=======	=======	=======	=======	||	=======
Marginal	0.21612	0.67347	0.11014	0.00027	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.74922	0.84126	0.90816	0.99973

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.09563	0.10538	0.001174	.000000048	||	0.2022
Partially Meets the Standards	0.10538	0.53149	0.060951	.000040352	||	0.6980
Meets the Standards	0.00117	0.06095	0.037430	.000113130	||	0.0997
Exceeds the Standards	0.00000	0.00004	0.000113	.000001341	||	0.0002
	=========	=========	=========	==========	||	========
	0.20218	0.69786	0.099668	.000154870		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.10222	0.10167	0.00130	.000000083	||	0.20520
Partially Meets the Standards	0.11264	0.51282	0.06735	.000069782	||	0.69300
Meets the Standards	0.00126	0.05882	0.04136	.000195622	||	0.10163
Exceeds the Standards	0.00000	0.00004	0.00012	.000002318	||	0.00017
	=========	=========	=========	==========	||	========
	0.21611	0.67334	0.11013	.000267805		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.65652	0.78312	0.87117	0.99957	||	0.28104

Accuracy and Consistency of Classifications
Grade 4 Mathematics

```
                         Step 4


                  Predicted Classification X(1)

                 Does Not   Partially                Exceeds
                 Meet the   Meets the   Meets the      the
tstat            Standards  Standards   Standards    Standards   ||   Marginal

Does Not Meet the Standards   0.22406    0.03737    0.00001   1.0759E-13   ||    0.26141
Partially Meets the Standards 0.07341    0.40656    0.05028   .000010125   ||    0.53015
Meets the Standards           0.00004    0.05098    0.14816   .006054878   ||    0.20526
Exceeds the Standards         0.00000    0.00000    0.00143   .001588345   ||    0.00301
                              =========  =========  =========  ==========      ========
                              0.29751    0.49490    0.19987   .007653348       0.99984


                         Step 5


                   Actual Classification X(0)

                 Does Not   Partially                Exceeds
                 Meet the   Meets the   Meets the      the
tstat            Standards  Standards   Standards    Standards   ||   Marginal

Does Not Meet the Standards   0.17286    0.04072    7.74E-6    123E-15    ||    0.21359
Partially Meets the Standards 0.05664    0.44293    0.05597    0.00001    ||    0.55555
Meets the Standards           0.00003    0.05554    0.16496    0.00692    ||    0.22745
Exceeds the Standards         231E-16    8.07E-7    0.00159    0.00182    ||    0.0034
                              =======    =======    =======    =======         =======
Marginal                      0.22953    0.53919    0.22252    0.00875    ||         1



              Accuracy    Cut #1    Cut #2    Cut #3
              0.78257    0.90261   0.88844    0.99148


                         Step 6


                          X(1)

                 Does Not   Partially                Exceeds
                 Meet the   Meets the   Meets the      the
tstat            Standards  Standards   Standards    Standards   ||   Marginal

Does Not Meet the Standards   0.22040    0.07628    0.00080   .000000063   ||    0.2975
Partially Meets the Standards 0.07628    0.34949    0.06891   .000217110   ||    0.4949
Meets the Standards           0.00080    0.06891    0.12456   .005624771   ||    0.1999
Exceeds the Standards         0.00000    0.00022    0.00562   .001811504   ||    0.0077
                              =========  =========  =========  ==========      ========
                              0.29747    0.49489    0.19989   .007653448       1.0000


                         Step 7


                          X(0)

                 Does Not   Partially                Exceeds
                 Meet the   Meets the   Meets the      the
tstat            Standards  Standards   Standards    Standards   ||   Marginal

Does Not Meet the Standards   0.17004    0.08310    0.00089   .000000072   ||    0.25406
Partially Meets the Standards 0.05885    0.38074    0.07672   .000248253   ||    0.51658
Meets the Standards           0.00062    0.07507    0.13864   .006432533   ||    0.22079
Exceeds the Standards         0.00000    0.00024    0.00626   .002071381   ||    0.00857
                              =========  =========  =========  ==========      ========
                              0.22951    0.53915    0.22251   .008752239       1.00000



          Consistency    Cut #1    Cut #2    Cut #3    ||     kappa
           0.69154      0.85652   0.84620   0.98682    ||    0.49757
```

Accuracy and Consistency of Classifications
Grade 4 Science

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.27362	0.05244	0.000000	7.5354E-19	||	0.32605
Partially Meets the Standards	0.08556	0.52930	0.024418	.000000149	||	0.63928
Meets the Standards	0.00000	0.01488	0.019627	.000010310	||	0.03452
Exceeds the Standards	0.00000	0.00000	0.000000	0	||	0.00000
	=========	=========	=========	==========	||	========
	0.35918	0.59661	0.044045	.000010458	||	0.99985

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.2116	0.05973	1.67E-7	144E-18	||	0.27133
Partially Meets the Standards	0.06615	0.60299	0.0225	0.00003	||	0.69167
Meets the Standards	4.55E-7	0.01695	0.01809	0.00197	||	0.037
Exceeds the Standards	0	0	0	0	||	0
	=======	=======	=======	=======	||	=======
Marginal	0.27776	0.67966	0.04059	0.00199	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.83268	0.87412	0.96052	0.99801

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.26331	0.09576	0.000095	5.4172E-11	||	0.3592
Partially Meets the Standards	0.09576	0.47333	0.027611	.000001903	||	0.5967
Meets the Standards	0.00010	0.02761	0.016327	.000008520	||	0.0440
Exceeds the Standards	0.00000	0.00000	0.000009	.000000034	||	0.0000
	=========	=========	=========	==========	||	========
	0.35917	0.59670	0.044041	.000010457	||	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.20361	0.10909	0.000088	.000000010	||	0.31280
Partially Meets the Standards	0.07405	0.53906	0.025444	.000362813	||	0.63899
Meets the Standards	0.00007	0.03145	0.015047	.001624584	||	0.04820
Exceeds the Standards	0.00000	0.00000	0.000008	.000006482	||	0.00002
	=========	=========	=========	==========	||	========
	0.27774	0.67960	0.040587	.001993890	||	1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.75780	0.81669	0.94258	0.99800	||	0.49210

Accuracy and Consistency of Classifications

Grade 4 Social Studies

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.10014	0.03123	0.00000	2.174E-15	||	0.13138
Partially Meets the Standards	0.06074	0.50464	0.06313	.000004207	||	0.62854
Meets the Standards	0.00001	0.05522	0.17413	.005414009	||	0.23477
Exceeds the Standards	0.00000	0.00000	0.00211	.003148556	||	0.00526
	=========	=========	=========	==========	||	========
	0.16089	0.59110	0.23937	.008566772		0.99995

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.08173	0.03019	2.95E-6	302E-17	||	0.11193
Partially Meets the Standards	0.04957	0.48779	0.07528	5.85E-6	||	0.61264
Meets the Standards	9.02E-6	0.05338	0.20762	0.00753	||	0.26854
Exceeds the Standards	384E-18	2.83E-7	0.00252	0.00438	||	0.00689
	=======	=======	=======	=======	||	=======
Marginal	0.13132	0.57136	0.28542	0.01191	||	1

	Accuracy	Cut #1	Cut #2	Cut #3
	0.78152	0.92022	0.87133	0.98995

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.09583	0.06459	0.00046	.000000008	||	0.1609
Partially Meets the Standards	0.06459	0.44470	0.08165	.000115052	||	0.5911
Meets the Standards	0.00046	0.08165	0.15167	.005576134	||	0.2394
Exceeds the Standards	0.00000	0.00012	0.00558	.002875805	||	0.0086
	=========	=========	=========	==========	||	========
	0.16088	0.59106	0.23936	.008566998		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.07820	0.06244	0.00055	0.000000	||	0.14121
Partially Meets the Standards	0.05272	0.42987	0.09735	0.000160	||	0.58012
Meets the Standards	0.00038	0.07892	0.18085	0.007751	||	0.26791
Exceeds the Standards	0.00000	0.00011	0.00665	0.003997	||	0.01076
	=========	=========	=========	=========	||	========
	0.13130	0.57134	0.28540	0.011907		1.00000

	Consistency	Cut #1	Cut #2	Cut #3	||	kappa
	0.69295	0.88390	0.82252	0.98533	||	0.46452

Accuracy and Consistency of Classifications
Grade 4 Health

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.000000	0.00000	0.00000	0.000000	||	0.00000
Partially Meets the Standards	0.015785	0.57263	0.11528	0.000025	||	0.70374
Meets the Standards	0.000039	0.06682	0.20184	0.009285	||	0.27795
Exceeds the Standards	0.000000	0.00000	0.00534	0.012819	||	0.01816
	=========	=========	=========	=========	||	========
	0.015824	0.63945	0.32246	0.022129			0.99985

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0	0	0	0	||	0
Partially Meets the Standards	0.02509	0.59422	0.09954	0.00004	||	0.71888
Meets the Standards	0.00006	0.06933	0.17428	0.01379	||	0.25746
Exceeds the Standards	115E-15	2.08E-6	0.00461	0.01904	||	0.02365
	=======	=======	=======	=======	||	=======
Marginal	0.02515	0.66356	0.27843	0.03287	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.78754	0.97485	0.83103	0.98156

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.000625	0.01395	0.00125	0.000000	||	0.0158
Partially Meets the Standards	0.013948	0.49774	0.12756	0.000296	||	0.6396
Meets the Standards	0.001250	0.12756	0.18323	0.010427	||	0.3225
Exceeds the Standards	0.000000	0.00030	0.01043	0.011408	||	0.0221
	=========	=========	=========	=========	||	========
	0.015823	0.63955	0.32247	0.022131			1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.000993	0.01447	0.00108	0.000000	||	0.01654
Partially Meets the Standards	0.022167	0.51636	0.11014	0.000439	||	0.64917
Meets the Standards	0.001986	0.13232	0.15820	0.015484	||	0.30803
Exceeds the Standards	0.000000	0.00031	0.00900	0.016941	||	0.02625
	=========	=========	=========	=========	||	========
	0.025146	0.66346	0.27842	0.032864			1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.69257	0.96029	0.75370	0.97477	||	0.36243

Accuracy and Consistency of Classifications
Grade 4 Visual and Performing Arts

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.18057	0.04801	0.00171	0.000006	||	0.23032
Partially Meets the Standards	0.14807	0.31055	0.09758	0.003781	||	0.55994
Meets the Standards	0.00493	0.07031	0.11337	0.020958	||	0.20956
Exceeds the Standards	0.00000	0.00000	0.00001	0.000014	||	0.00003
	=========	=========	=========	=========		========
	0.33357	0.42887	0.21268	0.024758		0.99985

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.12913	0.05551	0.00184	8.67E-6	||	0.1865
Partially Meets the Standards	0.10589	0.35911	0.10498	0.00561	||	0.57559
Meets the Standards	0.00352	0.0813	0.12196	0.03109	||	0.23788
Exceeds the Standards	4.42E-9	1.27E-6	0.00002	0.00002	||	0.00004
	=======	=======	=======	=======		=======
Marginal	0.23854	0.49593	0.2288	0.03672	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.61023	0.83323	0.80274	0.96328

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.19730	0.11528	0.02026	0.000758	||	0.3336
Partially Meets the Standards	0.11528	0.21356	0.09253	0.007516	||	0.4289
Meets the Standards	0.02026	0.09253	0.08664	0.013245	||	0.2127
Exceeds the Standards	0.00076	0.00752	0.01324	0.003241	||	0.0248
	=========	=========	=========	=========		========
	0.33360	0.42889	0.21268	0.024759		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.14105	0.13327	0.02180	0.001124	||	0.29730
Partially Meets the Standards	0.08243	0.24695	0.09953	0.011147	||	0.44008
Meets the Standards	0.01449	0.10698	0.09320	0.019642	||	0.23434
Exceeds the Standards	0.00054	0.00869	0.01425	0.004807	||	0.02829
	=========	=========	=========	=========		========
	0.23851	0.49589	0.22878	0.036719		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.48605	0.74632	0.73567	0.94460	||	0.21675

Accuracy and Consistency of Classifications
Grade 8 Reading

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.04811	0.02044	0.00000	0.000000	||	0.06854
Partially Meets the Standards	0.07602	0.49707	0.04123	0.000002	||	0.61426
Meets the Standards	0.00001	0.07024	0.20691	0.010382	||	0.28754
Exceeds the Standards	0.00000	0.00000	0.00882	0.020679	||	0.02951
	=========	=========	=========	=========		========
	0.12414	0.58775	0.25696	0.031063		0.99984

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.04412	0.01659	8.33E-7	999E-19	||	0.06071
Partially Meets the Standards	0.06972	0.40354	0.06425	6.01E-7	||	0.53751
Meets the Standards	0.00001	0.05702	0.32245	0.00286	||	0.38233
Exceeds the Standards	641E-18	1.19E-6	0.01375	0.00569	||	0.01945
	=======	=======	=======	=======		=======
Marginal	0.11384	0.47716	0.40045	0.00855	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.77580	0.91368	0.87872	0.98339

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.05830	0.06555	0.00029	0.000000	||	0.1242
Partially Meets the Standards	0.06555	0.44867	0.07335	0.000165	||	0.5878
Meets the Standards	0.00029	0.07335	0.17004	0.013281	||	0.2570
Exceeds the Standards	0.00000	0.00016	0.01328	0.017620	||	0.0311
	=========	=========	=========	=========		========
	0.12414	0.58773	0.25696	0.031066		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.05345	0.05322	0.00045	.000000001	||	0.10713
Partially Meets the Standards	0.06012	0.36420	0.11430	.000045307	||	0.53873
Meets the Standards	0.00027	0.05955	0.26495	.003655434	||	0.32846
Exceeds the Standards	0.00000	0.00013	0.02069	.004849434	||	0.02568
	=========	=========	=========	==========		========
	0.11384	0.47710	0.40040	.008550176		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.68754	0.88594	0.82523	0.97547	||	0.47836

Accuracy and Consistency of Classifications

Grade 8 Writing

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.000000	0.00000	0.00000	0	||	0.00000
Partially Meets the Standards	0.098648	0.52588	0.07184	.000034750	||	0.69641
Meets the Standards	0.000110	0.06047	0.23779	.005105019	||	0.30347
Exceeds the Standards	0.000000	0.00000	0.00000	0	||	0.00000
	=========	=========	=========	==========	||	========
	0.098758	0.58635	0.30963	.005139768			0.99988

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0	0	0	0	||	0
Partially Meets the Standards	0.10206	0.45083	0.09111	0.00002	||	0.64402
Meets the Standards	0.00011	0.05183	0.3016	0.00243	||	0.35598
Exceeds the Standards	0	0	0	0	||	0
	=======	=======	=======	=======	||	=======
Marginal	0.10218	0.50266	0.39272	0.00244	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.75243	0.89782	0.85692	0.99756

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.020168	0.07552	0.00307	.000000596	||	0.0988
Partially Meets the Standards	0.075516	0.41785	0.09242	.000587344	||	0.5864
Meets the Standards	0.003070	0.09242	0.20972	.004410744	||	0.3097
Exceeds the Standards	0.000001	0.00059	0.00441	.000141025	||	0.0051
	=========	=========	=========	==========	||	========
	0.098755	0.58637	0.30962	.005139709			1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.02086	0.06473	0.00389	.000000283	||	0.08950
Partially Meets the Standards	0.07813	0.35815	0.11720	.000279367	||	0.55383
Meets the Standards	0.00318	0.07921	0.26599	.002097607	||	0.35051
Exceeds the Standards	0.00000	0.00050	0.00559	.000067070	||	0.00617
	=========	=========	=========	==========	||	========
	0.10216	0.50259	0.39268	.002444327			1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.64514	0.85006	0.79571	0.99152	||	0.38264

Accuracy and Consistency of Classifications
Grade 8 Mathematics

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.35760	0.03419	0.00001	1.2852E-12	||	0.39185
Partially Meets the Standards	0.07925	0.30237	0.03647	.000016913	||	0.41809
Meets the Standards	0.00009	0.04805	0.13467	.007160187	||	0.18997
Exceeds the Standards	0.00000	0.00000	0.00000	0	||	0.00000
	=========	=========	=========	==========	||	========
	0.43695	0.38461	0.17115	.007177100		0.99991

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.28801	0.03948	0.00001	173E-14	||	0.32751
Partially Meets the Standards	0.06383	0.34912	0.04141	0.00002	||	0.45438
Meets the Standards	0.00007	0.05548	0.15291	0.00965	||	0.21812
Exceeds the Standards	0	0	0	0	||	0
	=======	=======	=======	=======	||	=======
Marginal	0.35191	0.44408	0.19433	0.00967	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.79005	0.89661	0.90301	0.99033

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.36060	0.07527	0.00108	.000000291	||	0.4370
Partially Meets the Standards	0.07527	0.25269	0.05626	.000375986	||	0.3847
Meets the Standards	0.00108	0.05626	0.10803	.005787849	||	0.1712
Exceeds the Standards	0.00000	0.00038	0.00579	.001013041	||	0.0072
	=========	=========	=========	==========	||	========
	0.43695	0.38459	0.17116	.007177167		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.29041	0.08690	0.00123	.000000392	||	0.37855
Partially Meets the Standards	0.06062	0.29175	0.06386	.000506639	||	0.41679
Meets the Standards	0.00087	0.06494	0.12265	.007800102	||	0.19629
Exceeds the Standards	0.00000	0.00043	0.00657	.001365185	||	0.00837
	=========	=========	=========	==========	||	========
	0.35190	0.44402	0.19431	.009672319		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.70622	0.85036	0.86813	0.98469	||	0.54344

Accuracy and Consistency of Classifications

Grade 8 Science

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.24951	0.04951	0.00003	8.0147E-13	||	0.29901
Partially Meets the Standards	0.09113	0.41400	0.05909	.000014745	||	0.56421
Meets the Standards	0.00004	0.02976	0.09776	.004034042	||	0.13159
Exceeds the Standards	0.00000	0.00000	0.00170	.003329754	||	0.00503
	=========	=========	=========	==========	||	========
	0.34068	0.49327	0.15858	.007378541	||	0.99984

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.21364	0.05395	0.00003	843E-15	||	0.26762
Partially Meets the Standards	0.07803	0.45117	0.06073	0.00002	||	0.58995
Meets the Standards	0.00003	0.03243	0.10048	0.00424	||	0.13718
Exceeds the Standards	114E-15	6.51E-7	0.00175	0.0035	||	0.00525
	=======	=======	=======	=======	||	=======
Marginal	0.29171	0.53755	0.16298	0.00776	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.76879	0.86795	0.90675	0.99400

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.24359	0.09555	0.00155	.000000123	||	0.3407
Partially Meets the Standards	0.09555	0.33630	0.06127	.000141233	||	0.4933
Meets the Standards	0.00155	0.06127	0.09157	.004191399	||	0.1586
Exceeds the Standards	0.00000	0.00014	0.00419	.003045559	||	0.0074
	=========	=========	=========	==========	||	========
	0.34069	0.49327	0.15858	.007378314	||	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.20856	0.10411	0.00159	.000000130	||	0.31428
Partially Meets the Standards	0.08180	0.36646	0.06297	.000148505	||	0.51144
Meets the Standards	0.00133	0.06676	0.09410	.004406929	||	0.16661
Exceeds the Standards	0.00000	0.00015	0.00431	.003201962	||	0.00766
	=========	=========	=========	==========	||	========
	0.29169	0.53748	0.16297	.007757525	||	1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.67239	0.81114	0.86704	0.99098	||	0.45954

Accuracy and Consistency of Classifications
Grade 8 Social Studies

```
                              Step 4


               Predicted Classification X(1)

              Does Not    Partially                Exceeds
              Meet the    Meets the    Meets the     the
tstat         Standards   Standards    Standards   Standards   ||   Marginal

Does Not Meet the Standards    0.18137     0.03541     0.00000     0.000000   ||   0.21680
Partially Meets the Standards  0.07860     0.44867     0.04111     0.000009   ||   0.56836
Meets the Standards            0.00003     0.05740     0.13715     0.006862   ||   0.20142
Exceeds the Standards          0.00000     0.00000     0.00405     0.009283   ||   0.01334
                               =========   =========   =========   =========       ========
                               0.25999     0.54148     0.18232     0.016153        0.99991


                              Step 5


               Actual Classification X(0)

              Does Not    Partially                Exceeds
              Meet the    Meets the    Meets the     the
tstat         Standards   Standards    Standards   Standards   ||   Marginal

Does Not Meet the Standards    0.13648     0.03532     4.09E-6     619E-16    ||   0.1718
Partially Meets the Standards  0.05915     0.44744     0.05574     9.27E-6    ||   0.56234
Meets the Standards            0.00002     0.05724     0.18593     0.00729    ||   0.25049
Exceeds the Standards          239E-16     2.43E-6     0.0055      0.00987    ||   0.01537
                               =======     =======     =======     =======         =======
Marginal                       0.19566     0.54        0.24717     0.01717    ||         1


              Accuracy     Cut #1      Cut #2      Cut #3
              0.77972      0.90550     0.88698     0.98720


                              Step 6


                        X(1)

              Does Not    Partially                Exceeds
              Meet the    Meets the    Meets the     the
tstat         Standards   Standards    Standards   Standards   ||   Marginal

Does Not Meet the Standards    0.18085     0.07860     0.00057     0.000000   ||   0.2600
Partially Meets the Standards  0.07860     0.39771     0.06488     0.000252   ||   0.5415
Meets the Standards            0.00057     0.06488     0.10916     0.007700   ||   0.1823
Exceeds the Standards          0.00000     0.00025     0.00770     0.008202   ||   0.0162
                               =========   =========   =========   =========       ========
                               0.26001     0.54144     0.18231     0.016154        1.0000


                              Step 7


                        X(0)

              Does Not    Partially                Exceeds
              Meet the    Meets the    Meets the     the
tstat         Standards   Standards    Standards   Standards   ||   Marginal

Does Not Meet the Standards    0.13608     0.07838     0.00077     0.000000   ||   0.21524
Partially Meets the Standards  0.05914     0.39661     0.08795     0.000268   ||   0.54403
Meets the Standards            0.00043     0.06470     0.14798     0.008184   ||   0.22132
Exceeds the Standards          0.00000     0.00025     0.01044     0.008717   ||   0.01941
                               =========   =========   =========   =========       ========
                               0.19564     0.53994     0.24714     0.017169        1.00000


          Consistency    Cut #1      Cut #2      Cut #3     ||    kappa
          0.68946        0.86127     0.84561     0.98086    ||    0.49014
```

Accuracy and Consistency of Classifications
Grade 8 Health

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.001790	0.00129	0.00000	3.4952E-17	||	0.00308
Partially Meets the Standards	0.023899	0.50195	0.11444	.000004279	||	0.64026
Meets the Standards	0.000014	0.08354	0.27057	.002387524	||	0.35657
Exceeds the Standards	0.000000	0.00000	0.00000	.000001665	||	0.00000
	=========	=========	=========	==========		========
	0.025702	0.58678	0.38501	.002393469		0.99991

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.00181	0.00131	1.88E-7	803E-19	||	0.00312
Partially Meets the Standards	0.02412	0.51091	0.11036	9.83E-6	||	0.6454
Meets the Standards	0.00001	0.08504	0.26095	0.00548	||	0.35148
Exceeds the Standards	363E-21	256E-12	2.35E-6	3.82E-6	||	6.17E-6
	=======	=======	=======	=======		=======
Marginal	0.02594	0.59726	0.37131	0.0055	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.77367	0.97456	0.80458	0.99451

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.005307	0.01989	0.00050	.000000006	||	0.0257
Partially Meets the Standards	0.019890	0.43207	0.13470	.000089496	||	0.5868
Meets the Standards	0.000503	0.13470	0.24768	.002158642	||	0.3851
Exceeds the Standards	0.000000	0.00009	0.00216	.000145108	||	0.0024
	=========	=========	=========	==========		========
	0.025700	0.58675	0.38505	.002393251		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.005356	0.02024	0.00049	.000000013	||	0.02609
Partially Meets the Standards	0.020069	0.43976	0.12988	.000205487	||	0.58999
Meets the Standards	0.000508	0.13708	0.23883	.004956245	||	0.38142
Exceeds the Standards	0.000000	0.00009	0.00208	.000333190	||	0.00251
	=========	=========	=========	==========		========
	0.025933	0.59718	0.37128	.005494935		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.68434	0.95869	0.73170	0.99266	||	0.37531

Accuracy and Consistency of Classifications
Grade 8 Visual and Performing Arts

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.20114	0.04655	0.00576	0.000021	||	0.25348
Partially Meets the Standards	0.14255	0.21442	0.13910	0.006097	||	0.50220
Meets the Standards	0.00777	0.04895	0.14194	0.035370	||	0.23404
Exceeds the Standards	0.00000	0.00018	0.00418	0.005850	||	0.01021
	=========	=========	=========	=========		========
	0.35147	0.31010	0.29097	0.047338		0.99992

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.16211	0.06188	0.00519	0.00002	||	0.22921
Partially Meets the Standards	0.11488	0.28504	0.12537	0.00545	||	0.53073
Meets the Standards	0.00626	0.06507	0.12792	0.03159	||	0.23084
Exceeds the Standards	3.87E-6	0.00023	0.00376	0.00523	||	0.00923
	=======	=======	=======	=======		=======
Marginal	0.28326	0.41222	0.26224	0.04228	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.58029	0.81176	0.79241	0.95894

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.21506	0.09613	0.03858	0.001694	||	0.3515
Partially Meets the Standards	0.09613	0.11607	0.08989	0.008011	||	0.3101
Meets the Standards	0.03858	0.08989	0.13611	0.026413	||	0.2910
Exceeds the Standards	0.00169	0.00801	0.02641	0.011219	||	0.0473
	=========	=========	=========	=========		========
	0.35146	0.31010	0.29099	0.047337		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.17331	0.12778	0.03477	0.001513	||	0.33740
Partially Meets the Standards	0.07747	0.15427	0.08099	0.007155	||	0.31993
Meets the Standards	0.03109	0.11948	0.12267	0.023590	||	0.29684
Exceeds the Standards	0.00136	0.01065	0.02380	0.010019	||	0.04584
	=========	=========	=========	=========		========
	0.28323	0.41217	0.26223	0.042278		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.46031	0.72600	0.71297	0.93192	||	0.22096

Accuracy and Consistency of Classifications
Grade 11 Reading

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.001619	0.00101	0.00000	0.000000	||	0.00263
Partially Meets the Standards	0.038040	0.47565	0.06308	0.000002	||	0.57678
Meets the Standards	0.000013	0.08067	0.28241	0.012508	||	0.37561
Exceeds the Standards	0.000000	0.00000	0.01129	0.033653	||	0.04494
	=========	=========	=========	=========		========
	0.039672	0.55734	0.35678	0.046164		0.99997

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.00227	0.00078	8.22E-8	461E-20	||	0.00304
Partially Meets the Standards	0.0533	0.36356	0.08815	8.84E-7	||	0.50501
Meets the Standards	0.00002	0.06166	0.39464	0.00538	||	0.4617
Exceeds the Standards	568E-18	6.31E-7	0.01578	0.01447	||	0.03025
	=======	=======	=======	=======		=======
Marginal	0.05559	0.426	0.49857	0.01985	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.77494	0.94591	0.85017	0.97884

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.009813	0.02950	0.00037	0.000000	||	0.0397
Partially Meets the Standards	0.029495	0.43005	0.09761	0.000137	||	0.5574
Meets the Standards	0.000365	0.09761	0.24228	0.016506	||	0.3568
Exceeds the Standards	0.000000	0.00014	0.01651	0.029518	||	0.0462
	=========	=========	=========	=========		========
	0.039674	0.55730	0.35676	0.046161		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.013748	0.02254	0.00051	0.000000	||	0.03681
Partially Meets the Standards	0.041321	0.32874	0.13641	0.000059	||	0.50654
Meets the Standards	0.000511	0.07460	0.33856	0.007097	||	0.42079
Exceeds the Standards	0.000000	0.00010	0.02307	0.012690	||	0.03586
	=========	=========	=========	=========		========
	0.055580	0.42598	0.49855	0.019846		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.69374	0.93511	0.78778	0.96967	||	0.46427

Accuracy and Consistency of Classifications
Grade 11 Writing

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.03045	0.01361	0.00002	0.000000	||	0.04408
Partially Meets the Standards	0.07112	0.47424	0.10649	0.000057	||	0.65198
Meets the Standards	0.00014	0.06242	0.21533	0.011311	||	0.28918
Exceeds the Standards	0.00000	0.00000	0.00474	0.009970	||	0.01471
	=========	=========	=========	=========		========
	0.10171	0.55027	0.32658	0.021337		0.99995

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.02818	0.01225	0.00003	161E-14	||	0.04046
Partially Meets the Standards	0.06584	0.42678	0.13039	0.00003	||	0.62304
Meets the Standards	0.00013	0.05617	0.26366	0.00571	||	0.32567
Exceeds the Standards	302E-14	2.66E-6	0.0058	0.00503	||	0.01084
	=======	=======	=======	=======		=======
Marginal	0.09415	0.49521	0.39987	0.01077	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.72366	0.92176	0.81325	0.98846

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.04013	0.05880	0.00279	0.000001	||	0.1017
Partially Meets the Standards	0.05880	0.37732	0.11366	0.000480	||	0.5503
Meets the Standards	0.00279	0.11366	0.19849	0.011642	||	0.3266
Exceeds the Standards	0.00000	0.00048	0.01164	0.009212	||	0.0213
	=========	=========	=========	=========		========
	0.10172	0.55026	0.32658	0.021336		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.037140	0.05291	0.00341	0.000000	||	0.09347
Partially Meets the Standards	0.054420	0.33954	0.13916	0.000242	||	0.53341
Meets the Standards	0.002579	0.10228	0.24301	0.005875	||	0.35378
Exceeds the Standards	0.000001	0.00043	0.01426	0.004649	||	0.01934
	=========	=========	=========	=========		========
	0.094140	0.49516	0.39984	0.010766		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.62441	0.88667	0.75188	0.97919	||	0.35838

Accuracy and Consistency of Classifications
Grade 11 Mathematics

```
                                  Step 4


                        Predicted Classification X(1)

                   Does Not    Partially                 Exceeds
                   Meet the    Meets the    Meets the      the
tstat              Standards   Standards    Standards    Standards    ||    Marginal

Does Not Meet the Standards    0.34454      0.04624      0.00001      2.4133E-14   ||    0.39081
Partially Meets the Standards  0.04816      0.31989      0.04071      .000005315   ||    0.40875
Meets the Standards            0.00001      0.04179      0.15375      .004807472   ||    0.20035
Exceeds the Standards          0.00000      0.00000      0.00000         0         ||    0.00000
                               =========    =========    =========    ==========        ========
                               0.39271      0.40791      0.19446      .004812787        0.99991


                                  Step 5


                        Actual Classification X(0)

                   Does Not    Partially                 Exceeds
                   Meet the    Meets the    Meets the      the
tstat              Standards   Standards    Standards    Standards    ||    Marginal

Does Not Meet the Standards    0.32936      0.04427      6.73E-6      484E-16      ||    0.37363
Partially Meets the Standards  0.04603      0.30621      0.04699      0.00001      ||    0.39924
Meets the Standards            0.00001      0.04         0.17747      0.00964      ||    0.22712
Exceeds the Standards             0            0            0            0         ||       0
                               =======      =======      =======      =======           =======
Marginal                       0.3754       0.39048      0.22447      0.00965      ||       1



                   Accuracy      Cut #1       Cut #2       Cut #3
                   0.81304      0.90968      0.91298      0.99035


                                  Step 6


                                  X(1)

                   Does Not    Partially                 Exceeds
                   Meet the    Meets the    Meets the      the
tstat              Standards   Standards    Standards    Standards    ||    Marginal

Does Not Meet the Standards    0.32642      0.06587      0.00041      .000000016   ||    0.3928
Partially Meets the Standards  0.06587      0.28485      0.05700      .000165731   ||    0.4079
Meets the Standards            0.00041      0.05700      0.13284      .004205704   ||    0.1945
Exceeds the Standards          0.00000      0.00017      0.00421      .000441611   ||    0.0048
                               =========    =========    =========    ==========        ========
                               0.39270      0.40789      0.19446      .004813062        1.0000


                                  Step 7


                                  X(0)

                   Does Not    Partially                 Exceeds
                   Meet the    Meets the    Meets the      the
tstat              Standards   Standards    Standards    Standards    ||    Marginal

Does Not Meet the Standards    0.31201      0.06305      0.00047      .000000033   ||    0.37557
Partially Meets the Standards  0.06296      0.27264      0.06578      .000332355   ||    0.40179
Meets the Standards            0.00039      0.05456      0.15335      .008432388   ||    0.21674
Exceeds the Standards          0.00000      0.00016      0.00485      .000885606   ||    0.00590
                               =========    =========    =========    ==========        ========
                               0.37536      0.39041      0.22446      .009650382        1.00000



                   Consistency   Cut #1       Cut #2       Cut #3    ||    kappa
                   0.73898      0.87312      0.87829      0.98622    ||    0.60052
```

Accuracy and Consistency of Classifications
Grade 11 Science

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.26001	0.04588	0.000000	3.8527E-20	||	0.30591
Partially Meets the Standards	0.10887	0.50525	0.026833	.000000054	||	0.64099
Meets the Standards	0.00000	0.01444	0.037827	.000340700	||	0.05261
Exceeds the Standards	-0.00000	0.00000	0.000162	.000274301	||	0.00044
	=========	=========	=========	==========		========
	0.36888	0.56557	0.064822	.000615055		0.99995

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.19482	0.052	2.53E-7	692E-22	||	0.24683
Partially Meets the Standards	0.08156	0.57273	0.03369	9.74E-8	||	0.68799
Meets the Standards	4.15E-7	0.01637	0.0475	0.00061	||	0.06449
Exceeds the Standards	-17E-21	1.23E-9	0.0002	0.00049	||	0.0007
	=======	=======	=======	=======		=======
Marginal	0.27638	0.64111	0.0814	0.0011	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.81555	0.86643	0.94993	0.99918

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.26489	0.10384	0.000131	2.052E-11	||	0.3689
Partially Meets the Standards	0.10384	0.43207	0.029659	.000002638	||	0.5656
Meets the Standards	0.00013	0.02966	0.034660	.000376105	||	0.0648
Exceeds the Standards	0.00000	0.00000	0.000376	.000236362	||	0.0006
	=========	=========	=========	==========		========
	0.36886	0.56557	0.064826	.000615105		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.19846	0.11771	0.000164	3.6842E-11	||	0.31636
Partially Meets the Standards	0.07779	0.48975	0.037239	.000004735	||	0.60483
Meets the Standards	0.00010	0.03362	0.043518	.000675082	||	0.07791
Exceeds the Standards	0.00000	0.00000	0.000472	.000424325	||	0.00090
	=========	=========	=========	==========		========
	0.27634	0.64107	0.081393	.001104143		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.73221	0.80423	0.92887	0.99884	||	0.48349

Accuracy and Consistency of Classifications
Grade 11 Social Studies

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.27863	0.04083	0.00006	0.000000	||	0.31952
Partially Meets the Standards	0.10181	0.34021	0.05047	0.000091	||	0.49255
Meets the Standards	0.00032	0.05202	0.12288	0.009226	||	0.18445
Exceeds the Standards	0.00000	0.00001	0.00161	0.001773	||	0.00339
	=========	=========	=========	=========		========
	0.38075	0.43307	0.17501	0.011090		0.99991

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.21185	0.04609	0.00007	161E-12	||	0.258
Partially Meets the Standards	0.07739	0.38392	0.0612	0.00008	||	0.52259
Meets the Standards	0.00024	0.05871	0.149	0.00797	||	0.21592
Exceeds the Standards	552E-13	8.18E-6	0.00195	0.00153	||	0.00349
	=======	=======	=======	=======		=======
Marginal	0.28948	0.48872	0.21222	0.00958	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.74629	0.87621	0.87969	0.99000

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.28455	0.09334	0.00290	0.000004	||	0.3808
Partially Meets the Standards	0.09334	0.27197	0.06689	0.000850	||	0.4331
Meets the Standards	0.00290	0.06689	0.09747	0.007751	||	0.1750
Exceeds the Standards	0.00000	0.00085	0.00775	0.002487	||	0.0111
	=========	=========	=========	=========		========
	0.38079	0.43305	0.17502	0.011090		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.21631	0.10532	0.00352	.000003161	||	0.32517
Partially Meets the Standards	0.07095	0.30688	0.08110	.000733733	||	0.45974
Meets the Standards	0.00220	0.07549	0.11818	.006692886	||	0.20258
Exceeds the Standards	0.00000	0.00096	0.00940	.002147198	||	0.01251
	=========	=========	=========	==========		========
	0.28947	0.48864	0.21219	.009576979		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.64360	0.81799	0.83598	0.98221	||	0.44145

Accuracy and Consistency of Classifications
Grade 11 Health

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.009018	0.00572	0.00000	6.6516E-17	||	0.01474
Partially Meets the Standards	0.039230	0.54346	0.08211	.000003033	||	0.66492
Meets the Standards	0.000010	0.08522	0.23306	.002008915	||	0.32031
Exceeds the Standards	0.000000	0.00000	0.00000	0	||	0.00000
	=========	=========	=========	==========	||	========
	0.048258	0.63440	0.31517	.002011948			0.99997

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.00766	0.00584	3.28E-7	291E-18	||	0.0135
Partially Meets the Standards	0.03333	0.55436	0.07895	0.00001	||	0.66666
Meets the Standards	8.48E-6	0.08692	0.22414	0.00878	||	0.31984
Exceeds the Standards	0	0	0	0	||	0
	=======	=======	=======	=======	||	=======
Marginal	0.041	0.64711	0.30309	0.0088	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.78615	0.96082	0.83411	0.99120

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.014534	0.03337	0.00036	.000000004	||	0.0483
Partially Meets the Standards	0.033371	0.48706	0.11398	.000098109	||	0.6345
Meets the Standards	0.000357	0.11398	0.19901	.001825809	||	0.3152
Exceeds the Standards	0.000000	0.00010	0.00183	.000088081	||	0.0020
	=========	=========	=========	==========	||	========
	0.048262	0.63451	0.31517	.002012003			1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.012346	0.03403	0.00034	.000000016	||	0.04673
Partially Meets the Standards	0.028351	0.49670	0.10960	.000428975	||	0.63511
Meets the Standards	0.000303	0.11624	0.19138	.007982254	||	0.31592
Exceeds the Standards	0.000000	0.00010	0.00176	.000385106	||	0.00224
	=========	=========	=========	==========	||	========
	0.041000	0.64708	0.30308	.008796350			1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.70085	0.93696	0.77297	0.98973	||	0.39113

Accuracy and Consistency of Classifications
Grade 11 Visual and Performing Arts

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.53821	0.10773	0.00275	0.000007	||	0.64868
Partially Meets the Standards	0.04190	0.05465	0.01509	0.000645	||	0.11229
Meets the Standards	0.00710	0.08485	0.12604	0.020893	||	0.23889
Exceeds the Standards	0.00000	0.00000	0.00000	0.000000	||	0.00000
	=========	=========	=========	=========		========
	0.58721	0.24723	0.14388	0.021545		0.99986

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.34993	0.16259	0.0045	2.87E-6	||	0.51703
Partially Meets the Standards	0.02724	0.08249	0.02473	0.00028	||	0.13473
Meets the Standards	0.00462	0.12807	0.20656	0.00899	||	0.34824
Exceeds the Standards	0	0	0	0	||	0
	=======	=======	=======	=======		=======
Marginal	0.38178	0.37315	0.2358	0.00927	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.63898	0.80105	0.83780	0.99073

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.46704	0.11029	0.00927	0.000665	||	0.5873
Partially Meets the Standards	0.11029	0.07649	0.05289	0.007537	||	0.2472
Meets the Standards	0.00927	0.05289	0.07033	0.011408	||	0.1439
Exceeds the Standards	0.00067	0.00754	0.01141	0.001935	||	0.0215
	=========	=========	=========	=========		========
	0.58726	0.24721	0.14390	0.021546		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.30359	0.16644	0.01518	.000286162	||	0.48557
Partially Meets the Standards	0.07169	0.11546	0.08667	.003242016	||	0.27709
Meets the Standards	0.00602	0.07983	0.11523	.004906654	||	0.20601
Exceeds the Standards	0.00043	0.01138	0.01869	.000832677	||	0.03133
	=========	=========	=========	==========		========
	0.38173	0.37312	0.23578	.009267509		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.53517	0.73991	0.79693	0.96106	||	0.29821

APPENDIX B

2000-01 SAMPLE REPORTS

Grade 4 Assessment
December 2000 Administration

STATE OF MAINE
DEPARTMENT OF EDUCATION
23 State House Station
Augusta, ME 04333

June 2001

J. Duke Albanese
COMMISSIONER

MAINE
EDUCATIONAL
ASSESSMENT

Dear Parents/ Guardians:

The Legislature approved Maine's *Learning Results* in May of 1997, giving all schools standards to measure student learning. Our goal is for all students in Maine to demonstrate that they meet all standards defined in the *Learning Results*. The Maine Educational Assessment (MEA), the state test which has been administered for the past 16 years in grades 4, 8 and 11, has been rewritten so it tests the challenging subject matter specified in the *Learning Results*. Your student was in the third group to take this revised test. In the past, the MEA reported individual student scores only in reading and writing (called English/Language Arts in the *Learning Results*) and in mathematics. The revised MEA also includes individual results in science and technology, and in social studies.

During December of 2000, students in grades 4, 8, and 11 were tested in reading and writing as part of the MEA. The test included a composition, multiple-choice questions, short-answer questions, and essay questions (known as "constructed-response"). The report on the reverse side of this letter provides you with important information about your student's performance on this part of the MEA, along with a summary of school,

district, and state results. Please keep in mind that your student's score measures learning over the past 3-4 years, not just the work of the past year.

In March 2001 your student participated in the second administration of the MEA in the subjects of mathematics, science and technology, and social studies. We anticipate that the results of the March testing will be available in September.

Staff at your school can provide further information about school and district results, and about your student's performance overall. The MEA is just one part of the comprehensive assessment system your school uses to measure student learning and school success. MEA results are used at the school, district, and state levels to improve teaching and learning.

Sincerely,

J. Duke Albanese
Commissioner

Information on Maine's *Learning Results*

- The *Learning Results* were developed in eight content areas by thousands of Maine citizens.

- The MEA was rewritten by hundreds of Maine teachers to align with the *Learning Results*.

- Setting MEA performance standards based on the quality of student work was completed by hundreds of Maine teachers and citizens.

- For a copy of Maine's *Learning Results* either call 624-6629 or find them on-line at http://janus.state.me.us/education/lres/homepage.htm

---

Performance Levels and Score Ranges

On this assessment, results are reported as four performance levels using scaled scores that range from 501 to 580. **The chart below describes the quality of student work:**

☐ **Exceeds the Standards   (561 to 580)**
The student's work demonstrates exemplary accomplishment of content knowledge, analysis, problem solving, and communication skills.

☐ **Meets the Standards   (541 to 560)**
The student's work demonstrates consistent accomplishment of content knowledge, analysis, problem solving, and communication skills.

☐ **Partially Meets the Standards   (521 to 540)**
The student's work demonstrates inconsistent accomplishment of content knowledge, analysis, problem solving, and communication skills.

☐ **Does Not Meet the Standards   (501 to 520)**
The student's work demonstrates limited command of content knowledge, analysis, problem solving, and communication skills.

Maine State MEA Summary Results for December 2000 Administration


	% of Students
Exceeds the Standards	25% / 25%
Meets the Standards	25% / 25%
Partially Meets the Standards	25% / 25%
Does Not Meet the Standards	25% / 25%

☐ ELA Writing   ■ ELA Reading
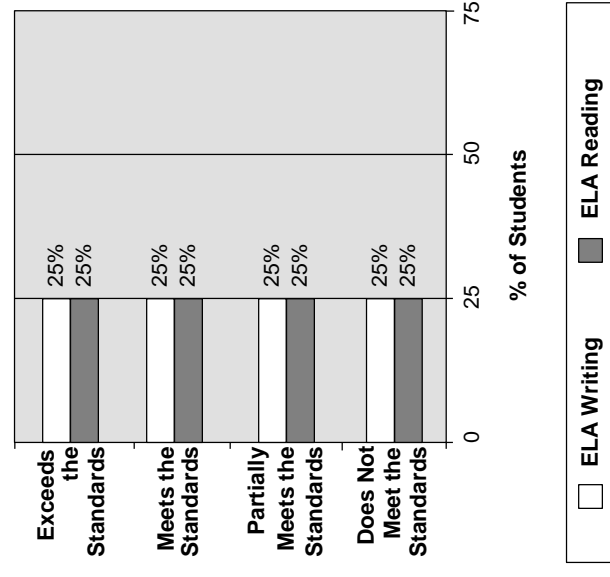
Student	Grade	School	District
	4		

This Student's Performance Levels and Scores

Content Area	Performance Level	Score	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards
ELA* Writing			◆			
ELA* Reading			◆			

*ELA is an abbreviation for English Language Arts.
Testing Incomplete (TI): Student failed to attempt one or more sessions.

501   520   540   560   580

See reverse side for description of performance levels and state summary results.

The ◆ represents the student's score. The bar (———) surrounding the score represents the probable range of scores for the student if he or she was tested many times. This statistic is called the standard error of measurement.

How this Student's Performance Compared to Average Scores from School, District, and State

ELA WRITING — Student 540, School 555, District 550, State 530
ELA READING — Student 540, School 555, District 550, State 530
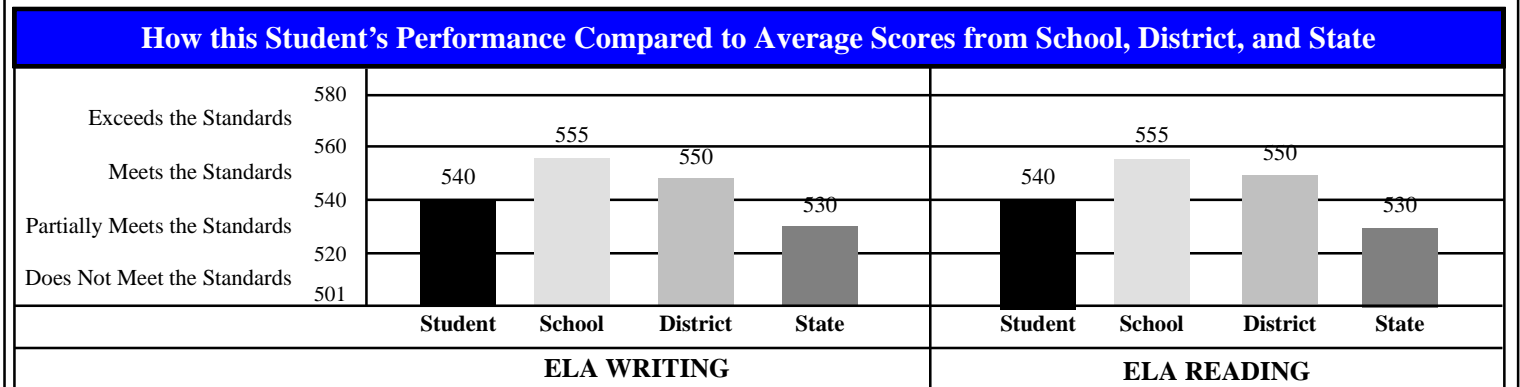
ELA Writing Comments

Commendations

The writer expressed creative or insightful ideas and/or demonstrated personal involvement.

The writer expressed creative or insightful ideas and/or demonstrated personal involvement.

The writer expressed creative or insightful ideas and/or demonstrated personal involvement.

The writer expressed creative or insightful ideas and/or demonstrated personal involvement.

Needs

The writer needed to demonstrate greater control of capitalization, spelling, and/or punctuation.

The writer needed more details and/or details that are more relevant to the topic.

The writer needed more details and/or details that are more relevant to the topic.

This Student's Performance in Content Area Subcategories

Content Areas	Content Area Subcategories	Student's Score Compared with Meeting the State Standards		
		Weaker	Meets the Standards	Stronger
ELA Writing	Standard English Conventions (Standard F)	◆		
ELA Writing	Stylistic and Rhetorical Aspects of Writing (Standard G)	◆		
ELA Reading	Reading Process, Language, and Comprehension (Standards A, B, C, D)	◆		

Definitions of Content Area Subcategories

Standard English Conventions: Refers to a student's ability to write correctly. Scoring focused on sentence structure, grammar and usage, and mechanics.

Stylistic and Rhetorical Aspects of Writing: Refers to a student's ability to use writing to explore ideas, to present lines of thought, to represent and reflect on human experience, and to communicate feelings, knowledge, and opinions. Scoring focused on topic development, organization, use of supportive details, and varied language and style.

Reading Process, Language, and Comprehension: Refers to a student's level of comprehension of literary reading selections (e.g., fiction, short stories, poetry) and informational reading selections (e.g., newspaper articles, informational essays, textbook passages), as well as a student's use of reading strategies, language, and analysis.

MAINE **E**DUCATIONAL **A**SSESSMENT

Name:
School:

District:
---------------- **Performance Levels** ----- **Scaled Scores**

Grade:　　**Writing:**
Date: 12/00　**Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* **required by law to be fully implemented by 2002-2003.**

MAINE **E**DUCATIONAL **A**SSESSMENT

Name:
School:

District:
---------------- **Performance Levels** ----- **Scaled Scores**

Grade:　　**Writing:**
Date: 12/00　**Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* **required by law to be fully implemented by 2002-2003.**

MAINE **E**DUCATIONAL **A**SSESSMENT

Name:
School:

District:
---------------- **Performance Levels** ----- **Scaled Scores**

Grade:　　**Writing:**
Date: 12/00　**Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* **required by law to be fully implemented by 2002-2003.**

MAINE **E**DUCATIONAL **A**SSESSMENT

Name:
School:

District:
---------------- **Performance Levels** ----- **Scaled Scores**

Grade:　　**Writing:**
Date: 12/00　**Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* **required by law to be fully implemented by 2002-2003.**

MAINE **E**DUCATIONAL **A**SSESSMENT

Name:
School:

District:
---------------- **Performance Levels** ----- **Scaled Scores**

Grade:　　**Writing:**
Date: 12/00　**Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* **required by law to be fully implemented by 2002-2003.**

MAINE **E**DUCATIONAL **A**SSESSMENT

Name:
School:

District:
---------------- **Performance Levels** ----- **Scaled Scores**

Grade:　　**Writing:**
Date: 12/00　**Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* **required by law to be fully implemented by 2002-2003.**

MAINE **E**DUCATIONAL **A**SSESSMENT

Name:
School:

District:
---------------- **Performance Levels** ----- **Scaled Scores**

Grade:　　**Writing:**
Date: 12/00　**Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* **required by law to be fully implemented by 2002-2003.**

MAINE **E**DUCATIONAL **A**SSESSMENT

Name:
School:

District:
---------------- **Performance Levels** ----- **Scaled Scores**

Grade:　　**Writing:**
Date: 12/00　**Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* **required by law to be fully implemented by 2002-2003.**

Educational Assessment Report

MAINE EDUCATIONAL ASSESSMENT

ID:

School:

District:

Grade: 4

Test Date: December 2000

Contents of the Report

The report is divided into five main sections including a section describing the students tested and a separate section for the results in each content area.

DEPARTMENT OF EDUCATION

2000-2001 School Year Reports

Dear School Board Members and School Personnel:

The Maine Educational Assessment (MEA) is in its third year of measuring student performance on Maine's *Learning Results*, which challenge schools and students to pursue academic standards that are among the highest in the nation. This report of student performance in reading, writing, and health education on tests administered in December 2000 is the first of two summary reports you will be receiving for the school year 2000 through 2001. The second report will include results from the assessment of mathematics, science and technology, social studies, and visual and performing arts administered in March of 2001. This MEA results report should still be considered baseline information, as the *Learning Results* are not scheduled for full implementation until the 2003 school year.

The MEA, revised to align with Maine's *Learning Results*, is composed of selected-response (multiple-choice) questions, as well as short-answer questions and complex questions, including a writing prompt, that require students to construct answers that demonstrate their knowledge and skills. Your review of the MEA questions that we have released will help you understand the revised assessment and the challenge that it presents for students and schools. The scores are reported using a numerical scale (501–580) and the performance levels "Does Not Meet the Standards," "Partially Meets the Standards," "Meets the Standards," and "Exceeds the Standards." The scale and the performance levels, established in the fall of 1999, will remain fixed for a period of at least five years to measure progress of students across the state in achieving the standards.

It is important to know that more than 500 teachers and other educators from across Maine helped to develop the revised MEA and assisted in the scoring and standard-setting process. Maine teachers continue to advise annually the updating and development of MEA tests. This participation has not only strengthened the redesigned MEA but has also engaged teachers from around the state in conversations about quality standards for student work.

I look forward to continuing the strong local and state partnership that has led to our current success as we work toward achieving even higher standards for all Maine students.

Sincerely,

J. Duke Albanese
Commissioner

Maine Educational Assessment

School:
District:
Grade: 4
Date: December 2000

SUMMARY OF SCORES

ELA WRITING

100% — 75% — 50% — 25% — 0%

	Exceeds			Meets			Partially Meets			Does Not Meet	
School	District	State	School	District	State	School	District	State	School	District	State

ELA READING

100% — 75% — 50% — 25% — 0%

	Exceeds			Meets			Partially Meets			Does Not Meet	
School	District	State	School	District	State	School	District	State	School	District	State

HEALTH EDUCATION

100% — 75% — 50% — 25% — 0%

	Exceeds			Meets			Partially Meets			Does Not Meet	
School	District	State	School	District	State	School	District	State	School	District	State

Executive Summary of School, District, and State Scores

Average Performance Score			
Year	School	District	State
ELA READING			
1998-1999			
1999-2000			
2000-2001			
Cum. Avg.			
ELA WRITING			
1998-1999			
1999-2000			
2000-2001			
Cum. Avg.			
HEALTH EDUCATION			
1998-1999			
1999-2000			
2000-2001			
Cum. Avg.			

MAINE EDUCATIONAL ASSESSMENT
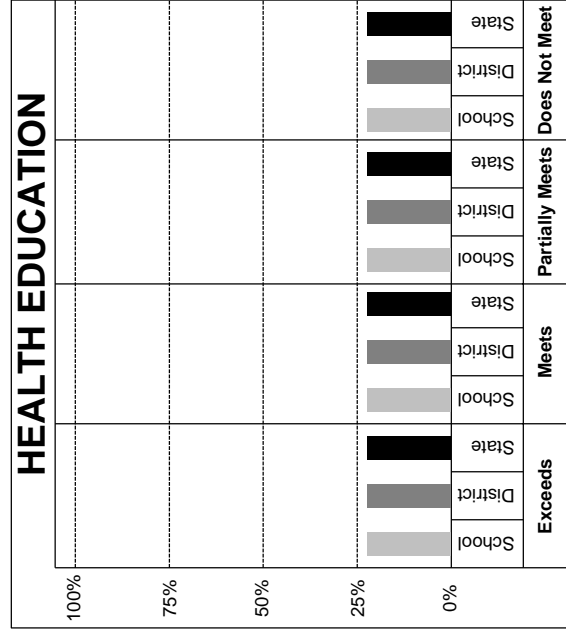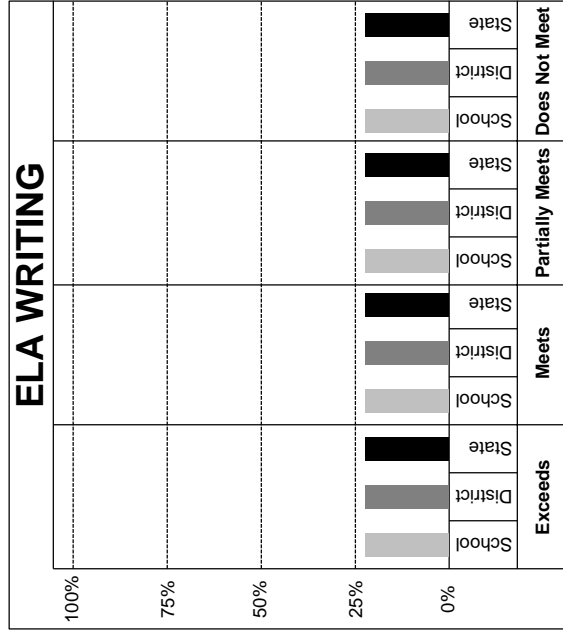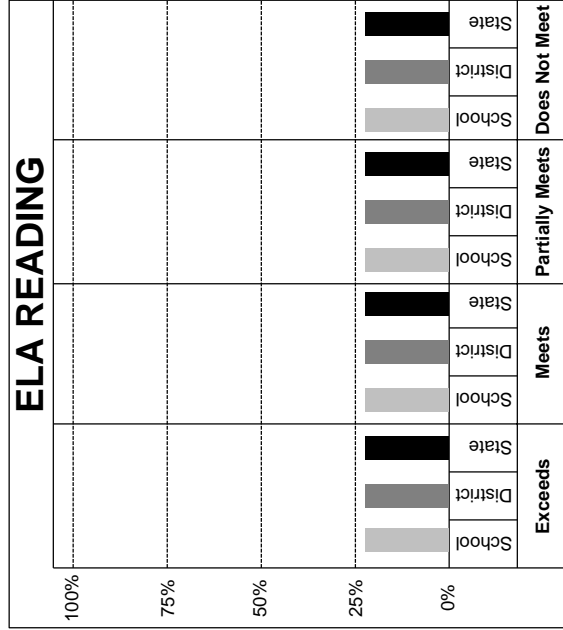
SUMMARY OF STUDENT PARTICIPATION

School:
District:
Grade: 4
Date: December 2000

Participation Category	Number			Percentage		
	State	District	School	State	District	School
Students Enrolled: number of completed test booklets						
Total Students Not Included in this Report:						
students who took no session of the assessment due to an identified disability						
students who took some but not all sessions of the assessment due to an identified disability						
students tested who receive special education and related services for more than 60% of the school day in a self-contained classroom as defined in Maine Special Education Regulations, Chapter 101, 11.6.						
students who took no session of the assessment due to LEP						
students who took some but not all sessions of the assessment due to LEP						
students who took no session of the assessment due to 504 Plan, absence, or other approved reason						
students who took some but not all sessions of the assessment due to 504 Plan, absence, or other approved reason						
Total Students Completing All Subjects:						
students with identified disability completing all subjects without accommodations						
students with identified disability completing all subjects with accommodations						
all others completing all subjects						
Percentage of Students with Identified Disability Included in Reports for All Subjects:						
Percentage of All Other Students Included in Reports for All Subjects:						

ELA READING RESULTS

MAINE EDUCATIONAL ASSESSMENT

School:
District:
Grade: 4
Date: December 2000

PERFORMANCE LEVELS

		STUDENTS AT EACH PERFORMANCE LEVEL			
		School		District	State
		N	%	N	%	%
Exceeds the Standards - The quality of a student's work at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The work demonstrates exemplary accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate. (Scaled scores: 561-580)	1998-1999				
	1999-2000				
	2000-2001				
	Cumulative Average				
Meets the Standards - The quality of a student's work at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The work demonstrates a consistent accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate. (Scaled scores: 541-560)	1998-1999				
	1999-2000				
	2000-2001				
	Cumulative Average				
Partially Meets the Standards - The quality of a student's work at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The work demonstrates inconsistent accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate. (Scaled scores: 521-540)	1998-1999				
	1999-2000				
	2000-2001				
	Cumulative Average				
Does Not Meet the Standards - The quality of a student's work at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The student demonstrates limited accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate. (Scaled scores: 501-520)	1998-1999				
	1999-2000				
	2000-2001				
	Cumulative Average				

Learning Results Content Standards	Average Points Attained (Number and Percent)					
	Number of Points Possible	School		District		State
		N	%	N	%	%
Reading Process and Language (Standards A and C)						
Reading Comprehension (Standards B and D)						
Literature & Culture (Standard B)						
Informational Texts (Standard D)						

Maine Educational Assessment

School:
District:
Grade: 4
Date: December 2000

Left Table: Reporting Categories

Reporting Categories	School				State				
	% Students in Each Category	% Exceeds the Standards	% Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards	% Students in Each Category	% Exceeds the Standards	% Meets the Standards	% Partially Meets the Standards / % Does Not Meet the Standards
GENDER									
girl									
boy									
GRADE FIRST ATTENDED SCHOOL IN DISTRICT									
kindergarten									
first or second grade									
third grade									
fourth grade									
TITLE 1 PROGRAM									
students currently served in reading									
students previously served in reading									
MIGRANT									
students eligible, not served									
students eligible, served, not tutored									
students eligible, served, tutored									
STATE-APPROVED GIFTED/ TALENTED PROGRAM									
yes									
no									
LANGUAGE MINORITY/LEP									
language minority student (bilingual never identified as LEP)									
former LEP student who has been reclassified non-LEP									
current LEP student									
IDENTIFIED DISABILITY									
yes									
no									
HOW MUCH TV DO YOU WATCH ON A SCHOOL NIGHT?									
none									
less than one hour									
one to two hours									
more than two hours									
OPTIONAL SCHOOL/DISTRICT QUESTION									
A									
B									
C									
D									
E									

Right Table: Questionnaire Items

Questionnaire Items	Sch. % Students in Each Category	State % Students in Each Category	% Exceeds the Standards	% Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards
How many pages do you read each day in school and to complete homework assignments?						
five or fewer pages						
six to ten pages						
eleven or more pages						
Do the questions on this MEA test reflect what you have learned in school about reading?						
Yes, the questions match the reading classes.						
They match somewhat.						
They matched just a little.						
There was no match.						
How many books have you read at home in the past two months?						
none						
one						
two to four						
five or more						
How often do you search for and read information on a computer?						
never						
once a month						
once a week						
two or more times a week						
How often do you do reading assignments or take tests where you earn points for what you have written even if your answer is not completely correct?						
never						
once a year						
once a week						
two or more times a week						
"In school I learn most of what I need to know to answer the MEA reading questions."						
It is true about me.						
It is not true about me.						
I am not sure.						
How good are you at reading?						
I am better than most students in my class.						
I am as good as most students in my class.						
I am not as good as most students in my class.						
How difficult were the reading sessions of the MEA test for you?						
very difficult						
difficult						
a little difficult						
not at all difficult						

ELA WRITING RESULTS

Maine Educational Assessment

School:
District:
Grade: 4
Date: December 2000

PERFORMANCE LEVELS

STUDENTS AT EACH PERFORMANCE LEVEL

PERFORMANCE LEVELS		School		District		State	
		N	%	N	%	N	%
Exceeds the Standards - The quality of a student's written compositions at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates exemplary accomplishment in both the development of the topic/idea and the use of standard English conventions in first draft writing. (Scaled scores: 561-580)	1998-1999 1999-2000 **2000-2001** Cumulative Average						
Meets the Standards - The quality of a student's written compositions at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates proficiency in both the development of the topic/idea and the use of standard English conventions in first draft writing. (Scaled scores: 541-560)	1998-1999 1999-2000 **2000-2001** Cumulative Average						
Partially Meets the Standards - The quality of a student's written compositions at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates writing skills that may show moderate development of the topic/idea and/or some errors in standard English conventions that may interfere with communication of ideas. (Scaled scores: 521-540)	1998-1999 1999-2000 **2000-2001** Cumulative Average						
Does Not Meet the Standards - The quality of a student's written compositions at this level does not meet the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates writing skills that show limited development of the topic/idea and/or many errors in standard English conventions that interfere with communication of ideas. (Scaled scores: 501-520)	1998-1999 1999-2000 **2000-2001** Cumulative Average						

Average Points Attained (Number and Percent)

Learning Results Content Standards	Number of Points Possible	School		District		State	
		N	%	N	%	N	%
Writing (Standards F and G)							
Standard English Conventions (Standard F)							
Stylistic and Rhetorical Aspects of Writing (Standard G)							

ELA WRITING RESULTS
(CONTINUED)

MAINE EDUCATIONAL ASSESSMENT

School:
District:
Grade: 4
Date: December 2000

Reporting Categories	School					State				
	% Students in Each Category	% Exceeds the Standards	% Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards	% Students in Each Category	% Exceeds the Standards	% Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards
GENDER										
girl										
boy										
GRADE FIRST ATTENDED SCHOOL IN DISTRICT										
kindergarten										
first or second grade										
third grade										
fourth grade										
TITLE 1 PROGRAM										
students currently served in reading										
students previously served in reading										
MIGRANT										
students eligible, not served										
students eligible, served, not tutored										
students eligible, served, tutored										
STATE-APPROVED GIFTED/TALENTED PROGRAM										
yes										
no										
LANGUAGE MINORITY/LEP										
language minority student (bilingual never identified as LEP)										
former LEP student who has been reclassified non-LEP										
current LEP student										
IDENTIFIED DISABILITY										
yes										
no										
HOW MUCH TV DO YOU WATCH ON A SCHOOL NIGHT?										
none										
less than one hour										
one to two hours										
more than two hours										
OPTIONAL SCHOOL/DISTRICT QUESTION										
A										
B										
C										
D										
E										

Questionnaire Items	Sch.	State				
	% Students in Each Category	% Students in Each Category	% Exceeds the Standards	% Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards
Do you or your teacher keep a collection of your writing?						
A collection of my writing is not kept.						
A collection of my writing is kept, but I do not use it.						
A collection of my writing is kept, and I use it to grow as a writer.						
How often do you have time in class to work on your writing?						
never						
a few times a week						
once a week						
almost every day						
How often does your teacher show you ways to improve/revise your writing?						
never						
a few times a month						
a few times a week						
almost every day						
How often does your teacher show you ways to edit your writing for spelling, capitalization, and punctuation?						
never						
a few times a month						
a few times a week						
almost every day						

ELA WRITING RESULTS
(CONTINUED)

School:
District:
Grade: 4
Date: December 2000

Summary of Annotations

ANALYTIC TRAITS	Number of Students			Percentage of Students		
	School	District	State	School	District	State
Topic Development Commendations Needs						
Organization Commendations Needs						
Details Commendations Needs						
Language/Style Commendations Needs						
Sentences Commendations Needs						
Grammar and Usage Commendations Needs						
Mechanics Commendations Needs						

The Annotated Holistic Scoring Method was used to evaluate students' papers. The two readers of each paper identified up to two evaluative statements or annotations addressing noteworthy traits of the paper. The annotations reflect either "commendations" or "needs" pertaining to the analytic traits of topic development, organization, details, language/style, sentences, grammar and usage, and mechanics.

This section reports the number and percentage of students in the school, district, and state receiving commendations or needs for each of the seven analytic traits. A student receiving more than one commendation or need for a particular element is counted only once in the determination of these data. The actual annotations appear on the Annotated Holistic Scoring Guide distributed with the student level results.

HEALTH EDUCATION RESULTS

School:
District:
Grade: 4
Date: December 2000

PERFORMANCE LEVELS

Exceeds the Standards - The quality of a student's body of work at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates exemplary knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis and risk reduction. (Scaled scores: 561-580)

Meets the Standards - The quality of a student's body of work at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates consistent knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis and risk reduction. (Scaled scores: 541-560)

Partially Meets the Standards - The quality of a student's body of work at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates partial and/or inconsistent knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis and risk reduction. (Scaled scores: 521-540)

Does Not Meet the Standards - The quality of a student's body of work at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates a limited knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis and risk reduction. (Scaled scores: 501-520)

STUDENTS AT EACH PERFORMANCE LEVEL

	School N	School %	District N	District %	State %
1998-1999					
1999-2000					
2000-2001					
Cumulative Average					
1998-1999					
1999-2000					
2000-2001					
Cumulative Average					
1998-1999					
1999-2000					
2000-2001					
Cumulative Average					
1998-1999					
1999-2000					
2000-2001					
Cumulative Average					

Average Points Attained (Number and Percent)

Learning Results Content Standards	Number of Points Possible	School N	School %	District N	District %	State N	State %
Health Concepts (Standard A)							
Health Information, Services, and Products (Standard B)							
Health Promotion and Risk Reduction (Standard C)							
Influences on Health (Standard D)							
Communication Skills (Standard E)							
Decision Making and Goal Setting (Standard F)							
Community, Consumer, and Environmental Health							
Personal and Nutritional Health							
Family Life Education and Growth and Development							
Safety and Injury Prevention							
Tobacco, Alcohol, and Other Drug Use Prevention							
Prevention and Control of Disease and Disorders							

HEALTH EDUCATION RESULTS
(CONTINUED)

MAINE EDUCATIONAL ASSESSMENT

School:
District:
Grade: 4
Date: December 2000

Reporting Categories

Reporting Categories	School				State			
	% Students in Each Category	% Exceeds the Standards	% Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards	% Exceeds the Standards	% Meets the Standards	% Partially Meets the Standards
GENDER girl boy								
GRADE FIRST ATTENDED SCHOOL IN DISTRICT kindergarten first or second grade third grade fourth grade								
TITLE 1 PROGRAM students currently served in reading students previously served in reading								
MIGRANT students eligible, not served students eligible, served, not tutored students eligible, served, tutored								
STATE-APPROVED GIFTED/ TALENTED PROGRAM yes no								
LANGUAGE MINORITY/LEP language minority student (bilingual never identified as LEP) former LEP student who has been reclassified non-LEP current LEP student								
IDENTIFIED DISABILITY yes no								
HOW MUCH TV DO YOU WATCH ON A SCHOOL NIGHT? none less than one hour one to two hours more that two hours								
OPTIONAL SCHOOL/DISTRICT QUESTION A B C D E								

Questionnaire Items

Questionnaire Items	Sch.	State				
	% Students in Each Category	% Exceeds the Standards	% Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards
How often do you have health education? almost every day once or twice a week once in a while never					
How much did you learn about nutrition this year? a lot some none					
How much did you learn about staying safe and preventing accidents this year? a lot some none					
How much did you learn about disease prevention this year? a lot some none					
How often do you do health assignments or take tests where you earn points for what you have written even if your answer is not completely correct? once a week once or twice a month once or twice a year never					
How well prepared do you feel you were to take the health test? very well prepared prepared not prepared at all I don't know.					
"In school I learn most of what I need to know to answer the MEA health questions." It is true about me. It is not true about me. I am not sure.					
How well can you learn the health that is taught in your school? very well well not too well not well at all					

Common Item Class Report

ELA READING
Grade 4

MAINE EDUCATIONAL ASSESSMENT

Code:
District:
School:
Class:
Date: December 2000
Group Size:
Page:

Item Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Points Earned (48 Max. Points)	Scaled Score	Performance Level
Content Standard & Performance Indicator	D3	D3	A5	A5	D3	A5	D3	C6	D4	A1	C6	B10	C6	B10	D3	A1	D1	D3	D4	B9	A1	C6	B6	C6	B9	B10	B6			
Item Type	MC	MC	MC	MC	MC	MC	SA	SA	CR	MC	MC	MC	SA	CR	MC	MC	MC	SA	CR	MC	MC	MC	MC	MC	MC	SA	ER			
Correct MC Response	D	C	D	A	B	C				A	C	D			A	B	A			D	B	A	B	B	C					
Total Possible Points	1	1	1	1	1	1	2	2	4	1	1	1	2	4	1	1	1	2	4	1	1	1	1	1	1	2	8			
Name																														

Item Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Percent Correct/Avg. Score: Class																											
Percent Correct/Avg. Score: School																											
Percent Correct/Avg. Score: District																											
Percent Correct/Avg. Score: State																											

Common Item Class Report

ELA WRITING

Grade 4

MAINE **E**DUCATIONAL **A**SSESSMENT

Code:
District:
School:
Class:
Date: **December 2000**
Group Size:
Page:

Name	Writing Prompt 20 Max. Points	Annotations	Extended-Response Question 10 Max. Points	Points Earned (30 Max. Points)	Scaled Score	Performance Level
Percent Correct/Avg. Score: Class						
Percent Correct/Avg. Score: School						
Percent Correct/Avg. Score: District						
Percent Correct/Avg. Score: State						